GEOGRAPHY AND MULTIDIMENSIONAL SCALING: ASSUMPTIONS, PROBLEMS AND APPLICATIONS

by

ARDEN CRAIG BRUMMELL, B.A.

Research Report

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Arts

McMaster University

April 1973

Acknowledgements

Any paper is the result of many contributors, never just one. This paper is no exception. I would especially like to thank Dr. John Betak, for his encouragement, numerous corrections and revisions, advice, and most of all, patience; and Elizabeth Harman, alias Libby, for the many, many, fruitful discussions and clarifications, which have so often disrupted our fellow neighbours, but without which this paper would not be possible. Finally, let us not forget the good wine, dear reader, which adds sparkle to the potations of this, or any other paper.

Arden C. Brummell
McMaster University
February, 1973.

Table of Contents

Geography and Multidimensional Scaling: Assumptions, Problems and Applications

I Introduction

This paper is concerned with multidimensional scaling concepts and techniques, and how these are applicable to geographic research. Basically, there are four purposes:

i) to examine the objectives of multidimensional scaling so as to understand the nature of the problem being considered and the types of questions being asked,

ii) to investigate the assumptions involved, (i.e., under what conditions is multidimensional scaling applicable?),

iii) to understand the conceptual problems in application and interpretation of MDS techniques,

iv) to examine applications in geography, providing a critical review of recent applications and projecting possibilities for future research.

The following section (II) attempts to demonstrate the relevance of psychological scaling (i.e., the assignment of numbers in a restricted manner to psychological phenomena), to geographic research. The geographic research of basic concern is that of "spatial behavior" or "individual spatial decision-making". Loosely, this may be termed behavioral geography. Section II is, in part, philosophical.

The paper then presents a general description of scaling concepts (III) and of the conceptualization of psychological scales as distances (IV). Similarities and preferential choice data are examined.

Three specific types of scaling routines are then examined (V, VI, VII), with special reference to assumptions, limitations and problems. The routines involved are related to 1) multidimensional scaling (MDS), involving stimuli alone, 2) multidimensional unfolding, (MDU), involving joint spaces of stimuli and individuals, and 3) INdividual Differences SCALing (INDSCAL), involving both stimulus spaces and individual spaces. MDS and INDSCAL use, primarily, similarities data; MDU uses preferential choice data. All three routines are nonmetric (or quasi-nonmetric in the case of INDSCAL) multidimensional scaling (NMS) techniques. Notice that the symbol, MDS, refers to both multidimensional techniques in general, and to the specific technique for scaling stimuli, listed as number one above.

Finally, some geographic applications of MDS and MDU are investigated. Some of the advantages and limitations of these techniques in a geographic context are explored. Some of the potentialities of INDSCAL are also suggested for geographic research (VIII).

II <u>The Relevance of Psychological Scaling to Geography</u>

Geographers have long been interested in explanations of spatial behavior, or more specifically, with the interaction of individuals with various points in the environment. Recently, several researchers have conceptualized this decision process as a choice problem, where an individual must select one spatial alternative, say, a grocery store, bank, or house, from a set of potential alternatives (Rushton, 1969a; 1969b; Ewing, 1970; Demko, 1971; Briggs, 1972). Clearly the decision is made on the basis of each individual's preferences for alternatives, which in turn are based on the individual's attitudes and perceptions (see for example, Gould, 1967; 1969; Rushton, 1969a). He chooses that alternative with the highest perceived or expected preference.[1]

This approach to spatial behavior reflects a desire, on the part of some geographers, for alternative explanations to spatial phenomena. Often, spatial models rely on initial postulates concerning individual behavior to derive an explanation of large scale pattern. These premises are assumed true, and tested only indirectly by the empirical tests of the derived theorems. It is well recognized, however, that these initial postulates are often "unrealistic". Yet, they remain because they permit the deduction of testable theorems which, in many cases, are strongly supported by the empirical test, and because more "realistic" postulates have not been developed. Many intuitively appealing observations about behavior are unsatisfactory as postulates, since they do not permit the logical deduction of meaningful theorems. More recently, however, advances in other areas, especially psychology, have permitted more direct analysis of the behavioral axioms. Instead of being limited to gross patterns of behavior, the geographer is now better able to examine the individual directly. The axioms of the collective models can now (hopefully) become the theorems of new models of individual behavior.

---

1. Notice, this does not assume optimizing behavior, but <u>intended</u> optimizing behavior. Secondly, the criterion of optimizing is preference, a personal multi-faceted phenomenon. Consequently, the individual is free to behave in seemingly irrational ways to the outside observer. Internally, we must assume consistency. at least in the short term, if we ever hope to discern behavioral laws.

Philosophically, this shift has two components. First, it represents a shift in the "level of focus", from the aggregate to the individual. Second, it represents a shift in the role the individual plays in explanation.

Essentially, this role-shift is a change in viewpoint from a behavioristic, stimulus-response approach[1], to a functional, cognitive approach which focuses on the internal activity of the individual (Chaplin & Krawiec, 1968). The cognitive approach assumes that individuals function in response, not to reality, but to their mental image of reality, and that they (individuals) adapt their behavior to changes in this perceived environment. Downs (1970b) has recorded this shift in philosophy in geography as a change from the traditional environment/behavior focus to an environment/man/behavior focus. In effect, the cognitive approach recognizes man as an intervening variable, both complex and interesting.

Of course, the purpose of this section is to show the relevance of psychological scaling to geographic problems. However, first, two positions need to be clarified, namely, the two philosophical shifts noted above.

First, scaling would seem relevant only if one is willing to adopt the philosophy that individuals respond to spatial stimuli <u>and</u> that the response, the behavior, is at least partially dependent on the individual's cognitive processes. That is, one must adopt a cognitive as opposed to a strict stimulus-response approach. Adopting this philosophy is, in effect, adopting the approach that individuals respond to their perceptions of reality and not necessarily reality itself. Therefore, we need to know about perceived attributes. Adopting this stance necessitates a connection between behavioral geography and psychological scaling. Without this stance, there is no connection.

Second, scaling would seem relevant only if one is willing to accept the usefulness of focusing on the individual as a basic unit of study. Psychological scaling is largely designed to examine data concerning individual behavior, not aggregate behavior (although in some cases it is useful for non-psychological data as well (Section VIII)). Consequently, if one veiws the geographic paradigm as being restricted to the description and explanation of large scale spatial structures, the usefulness of studying the individual, and hence, psychological scaling, may be questioned.

---

1. The behavioristic approach in psychology focuses on observables related to 1) characteristics of the stimuli, and 2) the nature of the response. The individual is viewed as a "black box" responding mechanistically to the stimuli. The nature of this mechanism is inferred from the observables, the stimuli and corresponding response (Chaplin & Krawiec, 1968).

There is a connection, of course. If individual behavior can be linked to aggregate behavior, then single unified theories can be developed. The problem of aggregation, however, is a sticky one and will be discussed a little later.

Accepting the cognitive approach and the usefulness of focusing on the individual, we now return to the choice situation. The problem is to determine the variables which influence preference. Long experience suggests that two sets of variables define spatial choices, or more simply, the desirability of any spatial alternative, 1) the attributes at a place, and 2) the relative spatial location of that place (Berry, 1964; Rushton, 1969a; 1969b; Briggs, 1972).

Geographic models have included variables falling into both these classes. The traditional central place models of Christaller (1933) and Loesch (1954), emphasize that interaction between the individual's residence and place of consumption, is inversely related to distance and directly related to the number and variety of functions at that place. Similarly, gravity-model formulations utilize the variables of distance and "mass", of which the latter at various times has included, size, income, employment, and number of telephones (see Olsson, 1965).

Briggs (1972, p. 1) focuses on these two sets of variables but emphasizes the cognitive approach:

> Recent urban research has stressed that spatial behavior is an outcome of the interaction between these two sets of environmental variables (attributes and relative location) and the cognitive processes of the individual. (parentheses mine)

Consequently, we might postulate that in any spatial choice or decision situation, preferences for alternative destinations[1] are a function of: a) the perceived attributes at a given point, and b) the perceived relative location of that point. By perceived attributes we mean those attributes which are recognized and meaningful for that individual. The problem is to divide attributes into those which are relevant and those which are irrelevant. There is no attempt to determine how the individual acquired this information, although this closely related problem has been considered by geographers (Golledge, 1969). By perceived relative location, we mean the cognized[2] distance from the individual to the destination, or the "closeness" of a given destination in relation to all other potential

---

1. Each destination presumably has some desirable feature for a given individual at a given time, even if it is the dentist's office (i.e., the place has some meaning to the individual).

2. In this paper, "perceived" is used synonymously with "cognized", meaning information which is stored, ordered, and retrievable by the individual.

destinations. Emphasis is placed on the relative closeness as distinct from absolute closeness.[1]

The broad problem of understanding behavior, then, is to define a preference function combining in some way, as yet unknown, relevant variables. The problem of determining which variables are relevant may be referred to as the problem of analysis; the problem of combining the variables into a preference function may be referred to as the problem of synthesis. It should be noted that both analysis and synthesis are context dependent. That is, stimuli must be comparable. At present, the problem of analysis is primary. Hopefully, as we gain more insight into the cognitive processes, the problem of synthesis can be pursued more fully.

Some attempts to 'analyze' variables have been made in geography. Lee (1970), Golledge, Briggs, and Demko (1969), Lowry (1970), and Briggs (1972), have attempted to define "cognitive distance". Similarly, Briggs (1969), Downs (1970a), Whipple and Niedell (1970), and Demko (1971) have attempted to define 'place' variables (attributes), for grocery stores, shopping centres, department stores, and cities, respectively.

Also, there have been attempts to 'synthesize' preference functions in geography (Rushton, 1969b; Ewing, 1970). Rushton assumes distance and size as relevant variables and uses these to define locational types. He then assumes preferences can be derived from aggregating actual individual choices. The locational type selected is clearly the most preferred of the potential types available, assuming the individual perceives all the "available opportunities". The choices are aggregated and converted into probabilities which are used as measures of similarity and input into an MDS routine to determine the relative strengths of preferences for locational types. This approach derives a preference function which is consistent with our intuitive expectations that preference increases with the number of opportunities (population) and decreases with travel costs(distance). It also determines the preferences from actual choices rather than potential choices of individuals. The method of recalling actual trips would seem less prone to 'reactive problems' (see Downs, 1970b), than the methods of introspective choice.

Rushton's research emphasizes the related problems of "relevant variables", "aggregation", and "level of focus". Rushton assumes that population and

---

1. Distance in this context refers to geographical distance.

distance are meaningful surrogates for more basic choice variables, and that the locational type classification is also meaningful. But we might well wonder if these are indeed meaningful to the individual household. This approach focuses on the individual (as a representative of the household), as the decision-maker. It is his behavior we are attempting to understand. Consequently, it seems important to examine the relevance of the variables to the individual. If we were concerned with a different level of focus, these variables might be extremely useful surrogates for explaining that structure. These two levels of focus can, of course, be linked if we assume, as Rushton does, that a group is merely a collection of independent individuals. Then the collective behavior can be obtained by aggregating the individual responses.

The reverse process can also occur, as we pointed out above. Postulates of individual behavior are often derived from observations of large scale phenomena. The postulates then become the axioms for deductive explanations of the collective pattern.

In either case, to explain geographic phenomena, there must be a wedding of the individual and the group. What are required are postulates of social behavior, probably more complex than simple aggregation. We now explore this problem.

Nagel (1961) has distinguished between "collective terms" and "individual terms". Often these collective terms, or concepts, are so poorly defined, that we cannot determine the variables involved, even to the extent of whether the variables themselves are collective or individual terms. Following, perhaps, a reductionist philosophy, many researchers have attempted to explain these collective systems by means of individual terms, that is, by inferring individual behavior from the aggregate or group behavior.

Arrow (1967), unfortunately, has aptly demonstrated that the choices of social actions made by a group depends on a constitution, welfare judgement, or valuation of the preferences of the individuals themselves. That is, a mere summation of individual choices will not necessarily reflect the choices of the group. Following Arrow (1967, p. 17), consider the following: A choice involves three alternatives, A, B, and C, one of which must be selected. One-third of the population has the preference ranking, A, B, C; one-third has the ranking, B, C, A; and one-third the ranking, C, A, B.

Considering only A and B, A will be preferred to B by a majority. Similarly, B will be preferred to C by a majority and C will be preferred to A by a majority. Clearly, the statements, A preferred to B, B preferred to C, and C preferred to A, are not transitive and hence logically contradictory.

So the question becomes: can we explain collective terms via individual terms? We cannot, except under certain restricted conditions (see Stokes, 1971), make sound inferences about individual terms from the aggregate. Can we deduce the collective terms from the individual terms? The answer is, not entirely. Nagel (1961, pp. 342-43) states the problem thus:

> ... no set of premises about the conduct of individual human beings might suffice for deducing some given statement about the actions of a group of men, and that at least one assumption of the latter kind might be required in any set of premises from which the given statement is deducible.

Essentially, deductions about groups require postulates about the social behavior of individuals.

This approach, however, from the individual to the group, seems to be fruitful in that if we can make realistic assumptions or postulates about individuals, then, perhaps, we can make assumptions about collective behavior which will allow us to deduce, and hence explain, the large scale phenomena (see also, Olsson, 1969).

But we might ask, how collective is the concept of, say, shopping. Clearly, it is made up of individuals. However, as Arrow points out, all preferences of individuals involve social actions, and it is impossible to determine those consequences of a social action which pertain only to the given individual.

In the shopping example, for instance, the individual is involved in a series of social behaviors. He is socially influenced by the opinions and expectations of friends and neighbours; he is socially constrained by streets and probably, the society-produced automobile; he is socially constrained by street signs, street lights, pedestrians, parking and other social laws; he is even socially influenced by language, definitely a social function. The store itself is a physical symbol of the social institution of shopping. The point to emphasize is that the individual is not isolated; his preferences for choices involve social choices which necessitate postulates about social interaction.[1] These social postulates are the crux of the matter.

---

1. Proshansky, Ittelson and Rivlin (1970) also emphasize the social context of all actions. Even the concept "isolation" has meaning only in relation to others.

Ideally, we should collect data at three levels --- the individual, the small group, and the aggregate. Destinctions between the individual and the small group may lead us to social postulates, which would permit deductions to the aggregate. However, while this approach would seem to be a valid strategy, we need not discard explanations dealing only with aggregates. Not all problems require analysis of individuals. Consequently, explanations at all levels would seem useful and desirable.

From this discussion, we might summarize this problem area in geography as: 1) determining the relevant or latent attributes perceived by an individual about a potential spatial alternative; distance, we might expect, will be one of the relevant attributes, and 2) determining some underlying rules of combination of these attributes to provide a more general and useful preference function. In short, we are concerned with the problems of analysis and synthesis.

MDS offers a solution to the problem of analysis, and provides numerical data for the problem of synthesis. MDS provides a means of forming a numerical scale which represents the relations among the perceived environmental phenomena, given certain assumptions. Phenomena are represented as points in a geometric space such that the distance between points coincides with the degree of similarity between the real world phenomena. The dimensions of the space are assumed to be the relevant attributes perceived by the individuals. Consequently, the derived space is known as a perceived attribute space or group stimulus space.[1] The scale values (i.e., coordinates), may then, with appropriate assumptions, be used to derive preference functions.

As with any technique, there are problems. A major problem is the question of how many dimensions are relevant. Also, how should the axes be oriented so that they can be interpreted. These, like many other problems in multidimensional formulations, are analagous to problems in factor analysis. It is the purpose of the remainder of the paper to make explicit the assumptions, conceptual problems, and application problems of MDS, both in general and in specific geographic problems. In particular, like factor analysis, multidimensional scaling implies an underlying psychological model which should be clearly understood for meaningful application.

It is hoped that the link between psychology and geography has been

---

1. See, for example, Coombs (1964); or Demko and Briggs (1970).

firmly established. This approach is not new, of course. Huff (1960), for example, was an early advocate of an inter-disciplinary approach. More recently, we can quote Harvey (1969, p. 95), mentally substituting 'geography' for 'economics':

> ... much of the work directed to defining and measuring abstract economic concepts such as 'utility', 'value', and so on, point to the general interest of economic research in seeking adequate definition by reference to psychological postulates.

III Concepts of Scaling

III.1 Definitions : Purposes of Scaling

In general, scaling is defined as: the actual process of assigning numbers to objects, or properties (Coombs, Dawes & Tversky, 1970, p. 31). It is distinguished from measurement which is concerned with the empirical conditions under which different types of scales can be constructed. Conceptualizing the measurement problem, however, may not lead to methods of scale construction.

We can subdivide scaling by distinguishing between "natural" physical attributes and psychological attributes. For example, consider an object such as a hammer, or an automobile. Both exhibit such readily measurable attributes (or properties) as width, length, weight, density, etc. However, these objects may also have important psychological attributes such as preference, meaning, quality, or intensity. Consider another example. A painting may be described by its length, width, andeven the wavelengths of its colours. Yet we expect it is some overall effect, a combination of psychological attributes which will elicit any behavioral response. In general, such attributes have no known measurable physical attributes. Psychological scaling, then, is that section of scaling in general, which is concerned with the assignment (and hence measurement) of numbers to these normally unquantifiable attributes.

The result of this scaling is the assignment of a number to each stimulus such that the number represents the magnitude of a particular attribute possessed by that stimulus. This quantification permits manipulation of the numbers by the logic of mathematics to derive results which can be interpreted for the empirical stimuli. More directly, scaling of psychological attributes permits the correlation of these attributes with 'physical' measurements of the stimuli and with behavioral responses to those stimuli (see Gullikson, "Forward", in Torgerson, 1958).

We can consider that this assignment of numbers, in effect, "measures" the attribute. It is clear that what is being measured is not the object, but the magnitude of the attribute possessed by the object. However, often the object itself is characterized by the number assigned as the value of the property. Often a stimulus is characterized by a set of numbers rather than a single number. This is the distinction between unidimensional and multidimensional scales.

Unidimensional scales measure the variation among objects on simply one attribute. A listing of the populations of a number of cities, the incomes of a group of people, or the preference rankings of a number of people for chocolate bars are examples of single attribute, unidimensional scales. But many objects are more complex than this and may be conceptualized as possessing several attributes. Chocolate bars, for example, may have the attributes of sweetness, texture and percentage of chocolate. Consequently, multidimensional scales measure the variation among objects on several attributes, so that each object is assigned a set of numbers. Geometrically, this is analogous to the difference between a line, say, the real line, R, and a plane, RxR, or a sphere, RxRxR, etc.: hence the concept of dimensions. For example, a shopping centre might be characterized by 1) the number of establishments, 2) the total floor area, 3) the total parking area, 4) the number of employees, 5) the quality of the lighting, and any other attributes one might think of as relevant to a given problem. The numbers assigned to each one of these attributes, then, would fix the location of that shopping centre on that particular dimension. The numbers related to the 'r' attributes would locate the centre in an 'r' dimensional space, so that each centre would be said to be characterized by an r-tuple of numbers.

III.2 Properties of Scales

We have defined scaling as the association of a set of empirical entities with numbers. This association, however, is not an ad-hoc rule of correspondence but a strict one to one, or isomorphic relationship. Each object is assigned only one number for each instance of a given property. It should be clear that there would be little purpose in scaling if the scale was not restricted to replicating, in some way, the order exhibited in the empirical objects. However, by order, we mean not that there is a particular number assigned to each object, but that the structure of relations between objects is fully represented by the numbers. Consequently, two scales using different rules, can assign different numbers to the same object for the same property, but maintain the order of relations. The degree that one scale can be varied but maintain the order of relations is termed its degree of uniqueness.

In the literature of measurement theory, these basic concepts are known as the representation problem and the uniqueness problem.[1] Briefly, the representation problem is the problem of assigning numbers to objects (i.e., scaling), in such a way that the relationships between the real world phenomena are fully represented by the abstract numbers. The concept of "fully represented" is embodied in the term "isomorphism" or "same structure" between the empirical and abstract relational systems. The uniqueness problem concerns the amount of variation, or tolerance, permitted in the assignment of each number in the scale. As the degree of uniqueness increases, the amount that each number can vary is reduced (i.e., more exact). Essentially, the degree of uniqueness depends on the nature of the properties of the empirical phenomena (i.e., physical or psychological), and the quality of the raw data. That is, the degree of uniqueness varies with the informational content available; as information increases, the scale becomes more unique. Finally, there is a third problem, that of meaningfulness. A scale is said to be meaningful if the truth value is invariant across all permissible transformations. In other words, a transformation permitted under a given degree of uniqueness should not affect the structural order (i.e., the relationships) between objects.[2]

IV Scales Conceptualized as Distances in Psychological Space

IV.1 Types of Scales

Scales can be classified not only on the basis of whether they analyze physical or psychological attributes, or whether they are one-dimensional or multidimensional, but also on the amount of information or degree of uniqueness which they possess. Normally, three types of scales --- ordinal, interval, and

---

1. A formal approach to these problems is presented by Suppes and Zinnes (1963); a less formal approach is clearly presented in Coombs, Dawes, and Tversky (1970).

2. For a fuller explanation of these measurement problems, particularly the uniqueness problem, see Appendix A.

ratio --- are distinguished.[1] These three classes are related to the important features of real numbers. Real numbers possess the following basic properties (distinct from operations which may be performed on the numbers), (Torgerson, 1958, p. 15):

1. Numbers are ordered.
2. Differences are ordered. That is, the interval between one pair of numbers is less than, equal to, or greater than the interval between another pair of numbers.
3. The series has a "zero" or unique origin.

These three properties coincide with the properties of ordinal, interval, and ratio scales. Torgerson (1958) terms them the properties of order, distance, and origin.

Order refers to a simple ranking (eg. 1, 2, 3, 4,...), involving the idea of inequality. For example, 2 is greater than 1; 4 is greater than 3; or x is greater than y. When we utilize variables such as x and y, we maintain the inequality but lose information concerning magnitude. Consequently, as long as the number associated with y is greater than the number associated with x, the order information or monotonicity requirement is maintained. The selection of actual numbers, however, has considerable freedom. For example, given $x < y$ (x less than y), we could choose, $x=1$, $y=2$; or $x=44$, $y=99$; or $x=6$, $y=402.397$. Each of these selections satisfies the order requirement and hence, are monotonic transformations of each other. They are order preserving. Thus, ordinal scales are said to be unique up to a monotonic transformation. Obviously, the order constraint has a low degree of uniqueness.

Distance (i.e., separation between two points) refers to ordered differences. That is, there is some information about the intervals between numbers. Given the information, $x > y > z$, $x-z > x-y > y-z$, we must choose numbers such that the interval xz is greater than xy, greater than yz. Letting $x=4$, $y=2$, $z=1$, satisfies all the distance (interval) and order requirements. Notice, $x=10$, $y=6$, $z=1$, does not satisfy the requirements, since $x-y < y-z$. Interval scales are said to be unique up to an increasing (positive) linear transformation, or one which preserves the order of differences. Notice, the origin is arbitrary.

Origin refers to the idea of a single, non-arbitrary origin. For example, given $x=4$, $y=2$, $z=0$, than an admissible alternative scale would be one which maintains the relation, $z=0$, and the ratio of the numbers x and y. One alternative scale might be, $x=8$, $y=4$, $z=0$. 'x' remains twice y and z remains at zero.

---

1. For other variations, see Appendix A.

This transformation can be performed simply by multiplying by a scalar (constant). Hence, ratio scales are said to be unique up to a scalar multiplication, or proportionality constant, or similarity transformation.[1]

In III.1, a distinction is made between physical properties or attributes such as length, width, weight, etc. which could be "naturally" measured, and psychological properties such as emotion, preference, quality, intensity, etc., which could not be so measured. This distinction has importnat implications in how measurement is conceptualized.

Psychological measurements are often conceived of as distances, whereas physical measurements are not. The distinction rests on the numerical properties of distance and origin. Physical attributes generally possess a natural origin; on the other hand, psychological attributes do not. For example, there is a natural, logical, "zero" weight for a pencil, or stone or automobile; but there is no such point for say, attitudes. Consequently, psychological measurements are more concerned with <u>differences</u> between pairs of points, which are conceived of as distances. The distance relation is of central importance. Physical measurement in contrast, is concerned with magnitudes from the origin, so that the operation of <u>addition</u> is of prime importance. Clearly, it is meaningful to add weights but not emotions. For this reason, psychological scales are often conceived of as a distance, which gives rise to an examination of the formal properties of a "distance function".

Before proceeding, however, it is important to distinguish between the terms "metric" and "nonmetric". Metric scales possess information about equality relationships, that is, the magnitude of differences, whereas, nonmetric scales possess information only about inequalities, that is, that one object is larger or smaller than another. Consequently, ratio and interval scales are (usually) metric measures, while ordinal scales are nonmetric measures (Niedell, 1969). Since the present formulations of multidimensional scaling focus on ordinal data as input, they are termed nonmetric multidimensional scaling (NMS) techniques. They produce, as output, however, metric information. For this reason, we shall now examine the distance or metric axioms.

IV.2 The Metric Axioms

The distance or metric axioms are commonly known as: 1) the property of

1. In mathematical terms ratio scales are unique up to a linear transformation or mapping. By comparison, interval scales are unique up to an affine transformation of the form, $X' = LX + Z$; X', X, Z refer to vectors, L is a linear operator.

positivity, 2) the property of symmetry, and 3) the triangle inequality (see Coombs, Dawes and Tversky, 1970, Appendix).

Consider a set X and a function d on X which assigns a unique number, d(x,y) to every pair (x,y) which is an element of XxX; then d(x,y) can be interpreted as "the distance between x and y".

1. Positivity (and Reflexitivity):

   $d(x,x) = 0$ That is, the distance between any point and itself is zero. Further, $d(x,y) > 0$ for all x, y, where $x \neq y$, which implies that distance is always nonnegative.

2. Symmetry:

   $d(x,y)=d(y,x)$ That is, that the distance from x to y is the same as from y to x; order is unimportant.

3. Triangle Inequality:

   $d(x,y) \leq d(y,z)+d(x,z)$ That is, considering a triangle, the sum of the distances of any two sides is greater than the third. Notice, if x,y,z, form a straight line, the equality condition results.[1]

A function which satisfies these properties is a distance function or metric. Notice that an ordinal scale does not satisfy these axioms, since, for example, there is no way of determining the truth of either the symmetry property or the triangle inequality.

In actual practice, distance, d(x,y), can be defined in a number of ways. Euclidean geometry defines distance as:

$$d(x,y) = (x^2+y^2)^{1/2}$$

This is the famous Pythagorean formula in two-dimensional space. However, Euclidean space is merely a special case of a family of metrics known as Minkowski-metrics or Lp norms. The definition of distance in the more general Minkowski formulation is the following:

$$d(x,y) = (\sum_{i=1}^{r} |x_i-y_i|^{p})^{1/p}$$

where x_i and y_i are the values of the points x and y respectively on the ith dimension. This function is basically a one parameter expression where p determines the weight assigned to each dimension.

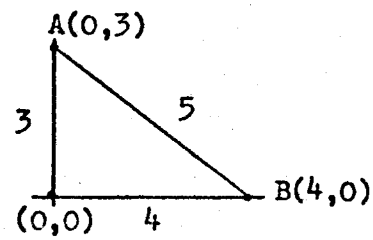
Notice, when p=2, we have the Euclidean case. We can see that in this case the distance between any two points is proportionally weighted to the differences along each axis. Or, that each component is proportionally considered in determining the distance.

When p=1, we have the "city block" or "Manhatten" metric. In this case,

1. The equality condition results if 1) the points form a straight line in Euclidean distance (Minkowski p=2), or 2) the distance is defined using the "city block"metric(Minkowski p=1).

the distance between two points is not measured along a straight line, but along the component line segments, so that the distance is the sum of the differences along each axis. For example, consider figure IV.1. In Euclidean space, A would be 5 units from B. With the 'city block' metric, A and B would be 7 units apart. Notice also that the x-axis contributes, in the Euclidean case, 16/25 ths of the total distance; but with the city-block metric, the x-axis contributes 4/7 ths of the total distance.

Figure IV.1:



Finally, when $p \to \infty$, we have the supremum (or dominance) metric. Then, the distance between two points is completely determined by the dimension with the largest difference. In figure IV.1, A would be 4 units from B, such that the x-axis would completely account for the distance between A and B.

The properties of Minkowski p-space have recently been examined in a penetrating paper by Beals, Krantz, and Tversky (1968). They point out three important properties:

1) Interdimensional Additivity: The distance between x and y is a function of the additive combination of the contributions of their components.

2) Intradimensional Subtractivity: The distance between x and y is a function of the absolute values of the component-wise differences.

3) Power: All component-wise differences are transformed by the same ($p \neq 1$) power function.

Notice that interdimensional additivity is concerned with additivity <u>across dimensions</u> so that the distance between two points is some "additive combination of component-wise effects", (Beals, Krantz & Tversky, 1968, p. 134). Intradimensional subtractivity is concerned with absolute differences along the <u>same dimension</u>. The power property defines how the component differences are to be transformed before their addition.

The reason for presenting these properties is to emphasize that the adoption of a particular metric implicitly implies a specific psychological model. Examination of the properties helps to clarify the assumptions which are being made. Clearly, it is important to recognize that when we use a Minkowski metric, we are implying a psychological model with the properties of additivity and subtractivity. Is a linear model appropriate for the way people think? Also, the selection of 'p', specifies how these components are to be combined to yield distance. What value should 'p' be?

Actually, with regard to 'p', there is little or no psychological support for any particular value in MDS formulations. Consequently, the familiar Euclidean metric has dominated research. Also, Euclidean space has the property that it is invariant under all rigid transformations, including rotation. No other Minkowski metric is invariant under rotation. Since, as we shall see, rotation is useful for interpretation, the Euclidean metric would seem to be the best choice in the absence of any conflicting hypotheses (see Demko and Briggs, 1971).

IV.3 Types of Data

The conceptualization of a scale as a distance has led Coombs (1964) to formulate a theory of data, based on the distance concept, in which he conceives of all psychological data as falling into classes distinguished by three dichotomies.

First, he distinguishes data on the basis of the type of relation between elements, that is, whether the relation is an order (dominance) relation or a proximity (consonance) relation.

Secondly, he distinguishes on the basis of whether the relation exists on a pair of points or on a pair of dyads (i.e., two pairs).

Thirdly, he distinguishes between comparisons drawn from the same set, and from distinctly different sets. For example, individuals and stimuli represent two different sets.

To clarify these distinctions we shall follow Coombs' conceptualization of "preferential choice data". It should be recognized at the onset, that Coombs is attempting to provide a general classification for all raw psychological data. He uses the basic idea that stimuli and/or individuals can be represented as points in a space so that the interpoint distances can be meaningfully interpreted. If we can conceptualize raw data as distances, then we are well on the way to achieving an interpretable metric scaling, which is the desired result.

Consider a set of objects, say, cars which we wish to order according to some attribute, say, power. We might envisage a line calibrated in some unit of power, say, horsepower, on which we can represent each car as a point. The difference between any two points can be interpreted as a distance.

Consider the situation in which several individuals are asked to rank these cars according to preference. Again we might envisage these stimuli represented by points on a line (or in a space), with the interpoint distances representing the differences in preference. Now, it may not be clear on what basis individuals determine preference. Some may compare cars simply on the strength of one attribute, say, power; others may consider several attributes, say, power, colour, styling, etc. Instead of locating each stimulus on a line, we might consider each stimulus being located in a space of several dimensions, where the perpendicular dimensions coincide with the relevant attribute dimension continuums perceived by each individual.

We might also consider representing the individuals as points in the same space, and interpreting the distance between any stimulus point and an individual's point as the degree of preference, so that points closer to the individual are more preferred. Consequently, the individual's point represents his ideal point.

Summarizing, we envisage a metric space, where the distance function is as yet undefined (some Minkowski metric), of 'r' dimensionality, where 'r' is to be determined experimentally to coincide with the latent attributes of the stimuli, in which both stimuli and individuals are represented by points. The preference by an individual for one stimulus over another is interpreted to mean that the preferred stimulus is nearer the individual's ideal point. Finally, the stimulus and ideal points are located in the space so as to reflect all the order relations, for all individuals, present in the observed relational system of preferences.[1]

This situation is one in which we are dealing with elements from two distinct sets. All behavioural data have something which assumes the active role and something which assumes the passive role. The stimulus can be any conceptual identity from a colour to a city. The active organism is most commonly an individual, but other imaginative interpretations are possible.[2] With preferential choice data, we are concerned with locating both sets, individuals and stimuli in a single psychological space. This is termed a joint space; both individuals and stimuli are jointly ordered. In contrast,

1. Notice that the actual location of the points is the scaling problem. The constraint is that all preference rankings are maintained in the point-space representation. That is, the order of stimuli away from an ideal point should match the ranking of stimuli by that individual.

2. For example, Rushton (1970) has located different types of stores as ideal points.

if we had been concerned simply with one set, usually stimuli, we would be dealing with a stimulus space.

It is evident that preferences demonstrate order relations. If, instead of asking individuals to rank the cars from most to least preferred, we asked them simply to indicate "yes" or "no", if they liked or disliked a car respectively, then we would have been eliciting a proximity relation. This can again be visualized as a distance. Consider a space in which individuals and cars are represented by points, and around each ideal point a distance a which varies from individual to individual. In two-dimensional space, this would be represented by a circle of radius a. This locus divides the space into two regions so that points "inside" can be interpreted as being close to the ideal and hence receiving a "yes", while points "outside" would receive a "no". This then, is a proximity relation.

It can be seen that preferential choice data consist of comparisons between two pairs of points, or dyads. In effect, we are comparing the distance from the individual's ideal point to one stimulus, with the distance from the ideal point to another stimulus. Each comparison involves two couples. In contrast, consider the situation in which individuals are asked to rank the cars according to some attribute, say, power. Then the comparison would be between pairs of stimuli; car A versus car B directly. In Coombs' terminology, these would be stimulus comparison data. Notice, that with this type of data, the specific attribute on which the stimuli are to be rated is explicitly given to the respondent. Also, all the responses are from the same set. On the other hand, if we wanted to rank both the stimuli and individuals (i.e., different sets) on given attributes, we would be considering single stimulus data. The essence of single stimulus data is that when an individual rates a stimulus he is also rating himself. Finally, if we ask individuals to compare dyads from the same set, then we are eliciting similarities data. For example, if we ask an individual to determine whether store A is more alike, or more similar, to store B than store C is to store D, then we will obtain an ordering of similarities AB and CD, as well as proximity information on the relative distances separating A and B, and C and D. If we have a common stimulus (i.e., a triad), we obtain information on the relative proximity of the two single stimuli from the common stimulus. Again, similarities data do not incorporate given attributes; the individual is free to make the judgements on his own personal bases.

The distinctions among the four types of data presented can be seen in figure IV.2. For our purposes, however, the most important types of data are

Figure IV.2 <u>The distinctions among Coombs' four basic types of data.</u>

	pairs of points	pairs of dyads
different sets	single stimulus	preferential choice
same set	stimulus comparison	similarities

preferential choice and similarities data. As we have seen, both these types of data involve judgements on dyads and can be conceptualized, or represented, in a general, abstract, manner, as a comparison of the distances between two pairs of points. These relations can be defined explicitly by the following expressions:

$$\left|P_{hij}\right| - \left|P_{hik}\right| \leq 0 \Leftrightarrow j \cdot\!> k$$

$$\left|P_{hi,jk}\right| - \left|P_{hi,j'k'}\right| \leq 0 \Leftrightarrow (j,k) < (j',k')$$

The first is concerned with preferential choice data and is interpreted as: at the moment h, the point corresponding to stimulus j is at least as preferred as stimulus k. P_{hij} represents the absolute distance between points i and j at moment h; P_{hik} represents the absolute distance between points i and k at moment h. If the value of the expression is negative, then k is farther away from i than is j. The symbol $\cdot\!>$ signifies "preferred to". The inclusion of a time factor indicates the realization that individuals may vary with time. The second expression is concerned with similarities data and is interpreted as: at the moment h, for individual i, the similarity between the stimulus j and the stimulus k is more than the similarity between stimuli j' and k'. Or, in distance terms, j is perceived closer or more alike k than j' is to k'.

What then, are the advantages of defining data in terms of distances. Immediately, there would seem to be two advantages. First, the concept of distance permits the use of the highly developed and powerful calculus of geometry to resolve the scaling problem. That is, one is not locating cars, or towns, or people, but points which can be manipulated more readily. Secondly, distance is a very general concept which permits many forms of raw data to be translated into distances. One of the reasons for the widespread use of MDS techniques is the ability to use a variety of data types. Shepard's (1962) initial paper used confusion data (the number of times A is confused with B is interpreted as a measure of similarity, and hence distance). Other kinds of data include, correlation coefficients, semantic differential scores, paired comparison judgements, factor scores, or any measure which indicates the degree of similarity or preference between two dyads.

We are now able to draw together some of the ideas in the last two sections. Essentially, we are interested in the way people comprehend phenomena. Specifically, we are interested in the dimensions and structure of peoples' cognitive images or field. That is, we would like to uncover the order in their psychological spaces. To uncover this order, however, we need to be able to measure and scale the dimensions of the cognitive structure. If we can determine the relative positions of a number of psychological phenomena, then it would seem that we would have some information for determining the structure itself. We then asked what we know about psychological phenomena. The most basic point was that psychological attributes have no unique "zero" point. Consequently, we need to concentrate on some sort of distance or interval-type scale. It is this degree of uniqueness which we can, at most, ever hope to obtain. We then explored some of the properties of distance or metric scales. These are the properties which we want our final scale to have. We then examined different types of psychological data. In particular, we were concerned with how we could reasonably convert or transform simple everyday questions, such as,"which car do you prefer?", into distance concepts. As we have seen, this conceptualization or interpretation has been performed so that many kinds or raw data can be interpreted as distances. Most important, this conceptualization demonstrates that locating stimuli and individuals as points in a space can lead to fruitful interpretation and meaning.

We know the properties that we would like our scale to possess. Also, we know that many types of data can be interpreted as distances; that is, the location of stimuli and individuals in space will have meaning. The problem then is, how do we convert information about distances to scale values in a space in which the distance properties will hold and the empirical relations (the raw distances) maintained? This is the problem we shall explore in the next three sections.

V Multidimensional Techniques: Multidimensional Scaling

The next three sections (V, VI, VII) are concerned with three different multidimensional scaling techniques. Multidimensional Scaling (MDS) uses similarities data to construct common group stimulus spaces; Multidimensional Unfolding (MDU) uses preferential choice data to construct joint spaces; and Individual Differences Scaling (INDSCAL) uses similarities data to construct both group and individual spaces. Under 'Multidimensional Unfolding' we shall also explore some newer, different models for constructing joint spaces, than the original 'Coombsian' unfolding model.

We now turn to an examination of Multidimensional Scaling (MDS) which is not to be confused with multidimensional techniques in general, although often the former is used to designate the entire range of techniques.

V.1 The Problem

Given a set of stimuli and a set of similarity measures on those stimuli, which vary with respect to a number of unknown attributes, determine a set of numbers for each stimulus so as to represent all relationships among the stimuli. The stimuli are conceived of as points in a metric space of 'r' dimensions, where 'r' is unknown, but represents the number of perceived attributes inherent in the stimuli, so that the projection of each point on an attribute axis represents the quantity of attribute which that stimulus possesses. Obviously, the coordinates of a point are the projections on the respective dimensions. The basic assumption of this interpretation of the calculus (the metric space) is that the similarities are a measure of interpoint distance in psychological space.

The problem, then, can be redefined as determining, 1) the dimensionality of the space, and 2) the relative location of the points in the space. The location of the points are constrained by the similarities measures which record the nearness of the other n-1 points (n= the number of stimuli). Hence, we are concerned with the relative location. Therefore, any zero point is purely arbitrary. The principle axes are unique only up to a positive linear transformation or, simply, an interval scale.

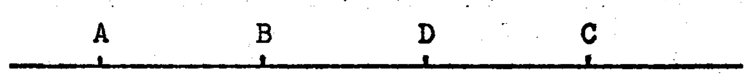
V.2 Conceptualization and Assumptions

Essentially, this formulation of the MDS problem makes two basic assumptions, 1) that the stimuli can be represented as points in a space, and 2) that interpoint distances between stimuli represent the degree of psychological similarity. As the interpoint distance declines, the degree of similarity increases.

In representing stimuli as points in a space, we would like a configuration such that the order of interpoint distances exactly matches the order of similarities, and for simplicity, we would like the dimensionality as small as possible. Obviously, the simplest case is a one dimensional scale. However, some similarities data cannot be accomodated by such a simple scale. Consider four stimuli, A, B, C, D, which have the similarities rankings, $BC < CD < AB < AD < BD < AC$. From the ordering, $AB < AD < AC$, we would expect the order on a one dimensional scale to be, A,B,D,C, as in figure V.1. However, we also know

Figure V.1

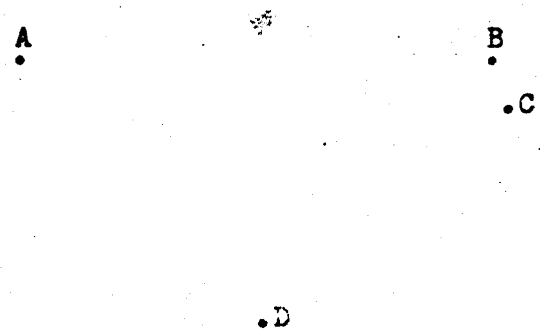
A one dimensional scale satisfying the ordering, $AB < AD < AC$.



that $BC < CD$, which is a contradiction in one dimension. But, if we allow a two-dimensional space, all inequalities can be satisfied, as in figure V.2.

Figure V.2

A two-dimensional scale satisfying the ordering, $BC < CD < AB < AD < BD < AC$. (Note: Axes may be included, but they are arbitrary. Only the order of distances are known.)



We might ask, however, what do these interpoint distances mean, or at least, how might they be interpreted? What do the coordinates mean? We might suspect that in judging the similarity of two stimuli, an individual is mentally locating each object on a set of perceived attributes, measuring the differences along each attribute, and then combining these differences in some way to achieve a single index of similarity. What the attributes are, and how they are weighted in the combination, are the problems of analysis and synthesis respectively.

If the mental process operates in this way, we might expect that the dimensionality of the psychological space will reflect the number of latent attributes which the individual considers. Consequently, we may interpret the components of the space as the subjective attributes of the stimuli. Now, we can interpret the interpoint distances as a reflection of some combination of these latent attributes.

Notice, however, that this interpretation makes two important assumptions. The first is that these data can be represented by a dimensionally organized space; hence, the dimensionality assumption. The second is that the psychological distance between stimuli can be represented by a metric distance; hence, the metric assumption (Beals, Krantz, and Tversky, 1968).

In general, we can list at least seven basic assumptions inherent in the MDS conceptualization:[1]

1) Individuals perceive stimuli as some combination of attributes.

2) Stimuli can be represented by points in a common metric space. (This is the metric assumption. See Beals, Krantz & Tversky, 1968)

3) The space is dimensionally organized so that the orthogonal dimensions represent attributes. (This is the dimensional assumption. See Beals, Krantz & Tversky, 1968)

4) Stimuli are located on the scale according to the effects of the attributes. In Minkowski space, the "effects" are an additive combination of the differences.

5) Similarities measures have inherent attribute measures so that the distances in the psychological space reflect the perceived similarities, or that, the interpoint distances are meaningful.

6) The resulting configuration is <u>valid for all individuals</u>, or that all individuals perceive the stimuli similarly. That is, individuals may vary and the number of attributes they use for comparison may vary, but the attributes they do use will be some subset of the total perceived attributes for those stimuli.

7) All stimuli are from the same set, that is, they are comparable.

---

1. See also, Demko, 1971, pp. 23-25.

V.3 Properties of MDS

Procedures now exist to solve the MDS problem in which, not only is a solution determined, but three important features are incorporated:[1]

1) Given the similarities between all pairs of stimuli, which are interpreted as interpoint distances in a space, the projections of the points on an arbitrary set of orthogonal axes can be determined.
2) Given ordinal, not metric information on the distances, a metric space can be recovered.
3) The choice of attributes for comparison is unspecified, leaving the individual a free comparison.

The first feature indicates that to solve the problem, that is, to provide an accurate multidimensional scale, we need, simply, a measure of interpoint distances. The second feature indicates that only an ordinal measure of these distances is required to generate a metric space (scale).

The second feature is extremely important. Individuals are seldom able to provide reliable information on an interval or ratio scale. The mental complexity is simply too great (see Simon, 1957). Efforts to elicit this type of information directly have generally led to anger and frustration in the respondents. In contrast, most individuals are able to provide ordinal information accurately and smoothly, without frustration.

The third feature indicates the unrestricted nature of these data. The individual is not confined to making comparisons about stimuli varying on prespecified attributes. With similarity comparisons, the respondent can choose as many attributes as he feels relevant. Hence, we obtain the location of an object in his psychological space and not some arbitrary attribute space.

The importance of this free approach can be stated another way. When the researcher forces the individual to make decisions on a prescribed set of attributes, any systematic variation related to unspecified, and hence unknown, attributes cannot be determined. This problem is avoided with the open approach.

A problem remains, however. While we feel more certain all variables are being considered in the free approach, we have no specific information on any one attribute. Consequently, there is a problem of interpreting the final configuration, and particularly "naming" the dimensions of the space. Exogenous information is required.

But, we might add further that it is the purpose of scaling to provide a series of numbers, characterizing the psychological relations among a set of objects so that these numbers may be correlated with physical phenomena in

---

1. See also, Golledge and Rushton, 1970.

an effort to develop theoretical interpretations. Although interpretation may be the ultimate objective, the purpose of scaling does not include interpretation. Scaling provides the information for subsequent interpretive analysis --- making inferences from the configuration --- which is a following phase. Again, the analogous position of factor analysis is emphasized.

V.4 The Historical Development of Multidimensional Scaling

V.41 The Metric Approach

Prior to the development of nonmetric iterative techniques, Torgerson (1958, Chapter 11) developed a multidimensional scaling procedure using metric (interval or ratio) data as input.

With this approach there are several problems. The first is to determine distances. One method is to use paired comparisons; convert the judgements into similarity proportions so that standard 'z' scores can be determined. These can then be interpreted as distances on an interval scale.

The second problem is to locate the points in a space of proper dimensionality. Given a dimensional Euclidean space, interpoint distances are easily determined. The reverse, however, is not so easy. Torgerson rationalized thus. First, assume that no error is involved. Consider three points, i, j, k, in a space. Then the distance between the points j and k can be defined as the scalar product (dot product, inner product) of the vectors $\overline{ij}$ and $\overline{ik}$. That is,

$$d(j,k) = d_{ij} \times d_{ik} \times \cos \theta_{jik}$$

where $d(j,k)$ is the distance between j and k, d_{ij} is the distance between i and j, d_{ik} is the distance between i and k, and $\cos \theta_{jik}$ is the cosine of the angle formed by the three points. Selecting one of the 'n' points in the space as an origin, a scalar product matrix, $\underline{B}$, can be defined. If error is involved, Torgerson suggests that the centroid of the points will be a better origin, since on average, the error involved with each point will cancel out. He presents a short-cut procedure for this translation.

Then, the scalar product matrix, $\underline{B}$, can be factored to obtain a matrix $\underline{A}$, where,

$$\underline{B} = \underline{AA'}$$

The factor scores are used as the coordinates of the stimuli in the space. The dimensionality of the space can be determined from the number of positive latent roots of the $\underline{B}$ matrix.

This yields a solution but has at least one major difficulty. This is the need for interval data, which requires an a priori distance function to translate the raw data to distances. Usually a linear transformation is assumed for simplicity. This however, precludes the possibility of a curvilinear relation. This is indeed a strong assumption, which need not be made with a nonmetric approach, and less than satisfactory without excellent data.

V.42 The Nonmetric Approach

The nonmetric approach is distinct from the metric approach in basically two ways, 1) the use of the monotonicity or order criterion, and 2) the use of an iterative process.

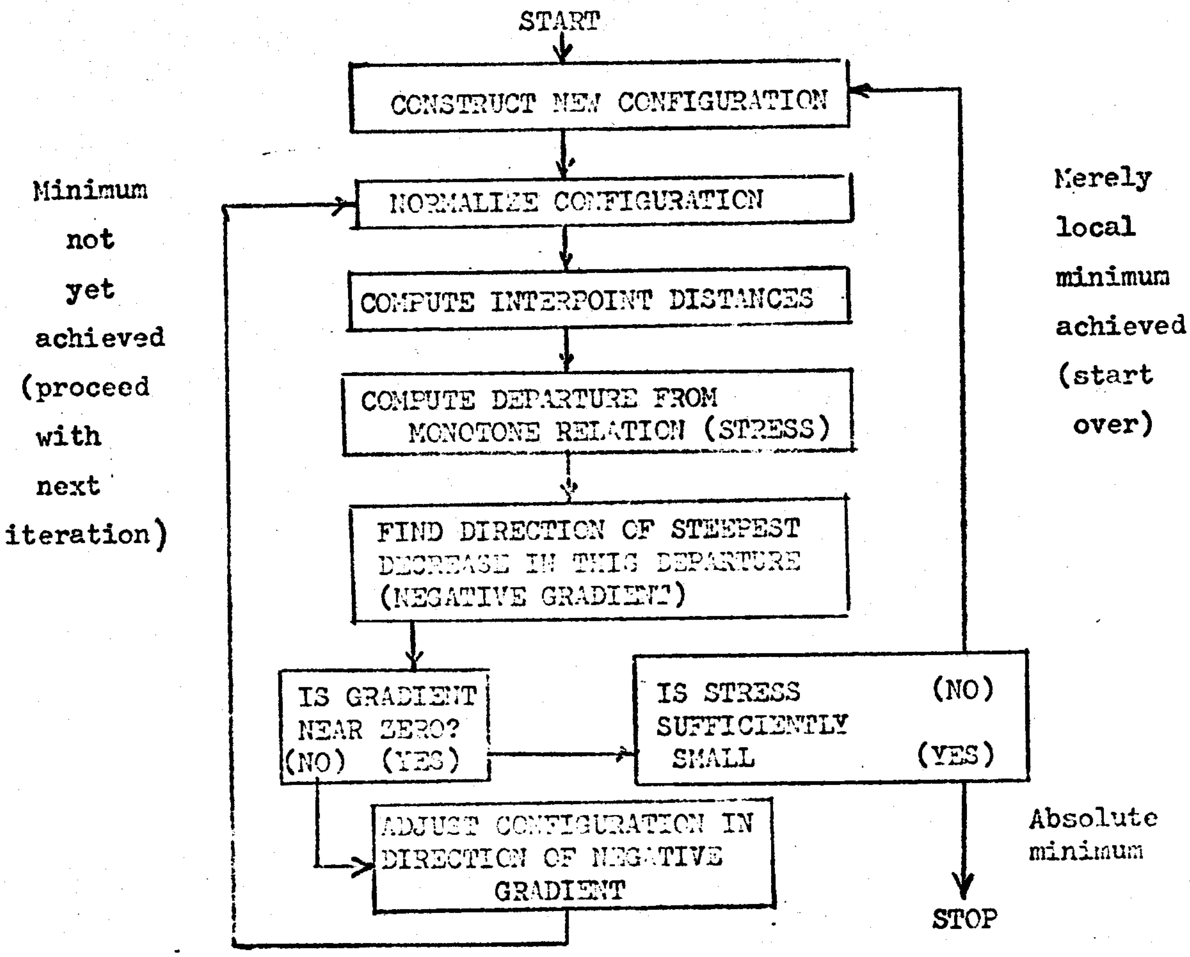
Both concepts were initiated by Shepard (1962) in an effort to derive a scale in which the order of the similarities were matched as closely as possible by the order of the interpoint distances, which represent the similarities. This is the monotonicity constraint. The metric approach, in contrast, requires an a priori selection of a specific function, or mapping, from the similarities to the necessary distance measures (i.e., transformed into a ratio or interval scale). In fact, Shepard first developed the monotonic approach in an effort to examine the nature of this distance function (exponential, straight-line, quadratic, etc.). His objective was to find a configuration of points so that he could plot psychological distances against the "real" distances, or similarities, thus revealing the shape of the transformation.

To solve the problem of locating the points in a space, Shepard developed an iterative procedure. This procedure may be described intuitively as follows. One begins by locating the points arbitrarily in a metric space, then computing the interpoint distances and comparing their rank order with the rank order of the initial similarities. (Shepard assumed, like Torgerson, that the similarities possessed a metric which was meaningful. Therefore, he maintained the similarities in their raw form.) Then, adjust each point a small amount so as to improve the monotonic relationship. Keep repeating until the monotonicity constraint (Kruskal uses a statistic called'stress') is as low as desired. A flow diagram of this type of iterative procedure is presented in figure V.3.

Consequently, at each iteration, all points are simultaneously displaced. How this displacement occurs is defined by two sets of n-1 vectors directed to the other n-1 points. One set indicates how the point should be moved so as to improve the monotonicity, that is, whether the point should be moved closer to, or farther from, each of the other n-1 points. The second set of n-1 vectors

Figure V.3

Flow diagram of iterative procedure (Kruskal Procedure).



Source: R. N. Shepard, "Introduction to Multidimensional Scaling Workshop", unpublished working paper prepared for Multidimensional Scaling Workship, June, 1972, University of Pennsylvania, p. 11.

aims at increasing the variation of distances, that is, encouraging large distances to become larger and small distances smaller. The purpose of this is to reduce the dimensionality of the configuration. Increasing the variation has the effect of stretching the configuration so that, for example, in two-space, a circle becomes more and more ellipitical and, at the limit, approaches a one-dimensional line. The resolution of the two vectors determines the movement of the point. The iterative procedure is halted when the statistic 'S' is small enough. Shepard defines 'S' as:

$$S = \left(\frac{2 \sum_{ij} (S_{ij} - S(d_{ij}))^2}{n(n-1)} \right)^{1/2}$$

where n = number of stimuli,

S_{ij} = a proximity measure between S_i and S_j,

$S(d_{ij})$ = a proximity measure related to the computed distance.

It is important to recognize one major problem with this statistic. It involves computations with similarities, the S_{ij}'s, and distances. This stems from Shepard's assumption that similarities have a meaningful extrinsic metric so that they can be manipulated in a raw (or modified) form. One of the major contributions of Kruskal (1964) was the recognition that we need not deal with the S_{ij}'s directly, but merely the rank order of the similarities. That is, the S_{ij}'s are transformed into positive natural numbers.

Finally, Shepard was left with n points located in n-1 dimensions. To reduce the dimensionality, he had to determine a scalar product matrix, factor, and discard "extra" factors.

Shepard's contribution should not be underemphasized, however. Most important, he emphasized, 1) the significance of the monotonicity requirement, and 2) that with this requirement, the configuration of points could be derived with an iterative procedure. Also, he demonstrated the metric nature of the configuration by comparing the iterative results derived from ordinal data concerning a known metric configuration, to that configuration, with excellent results.

Recent work by Shepard (1966) has more fully demonstrated that, in fact, order constraints do lead to metric-like scales. He states (p. 288):

> Actually though, if nonmetric constraints are imposed in sufficient number, they begin to act like metric constraints. In the case of a purely ordinal scale, the nonmetric constraints are relatively few and, consequently, the points on the scale can be moved about quite extensively without violating the inequalities (i.e., without interchanging any two points). As these same points are forced to satisfy more and more inequalities on the interpoint distances as well, however, the spacing tightens up until any but very small perturbations of the points will usually violate one or more of the inequalities.

The concept can also be stated another way. For 'n' points, there are n(n-1)/2 interpoint similarities measures, or order constraints. That is, for each point there are n-1 measures to the other n-1 points, or n-1 constraints on the location of each point. And, as the number of points increase, the number of constraints on the configuration increases not by a factor of n , but by a factor of n(n-1), or close to n^2. Therefore, as n increases arithmetically, the number of constraints increases geometrically, increasing the rigidity of the structure. This insight gives us greater confidence in the assumption that the derived configuration is, indeed, embedded in a metric space.

Shepard's emphasis on monotone relations rather than specific distance functions, prompted the work of Kruskal (1964a; 1964b). Kruskal, however, altered the problem somewhat, by defining, in effect, a more general problem. Instead of dealing with similarities directly, he derived a configuration in which the interpoint distances would be related monotonically to a simple ranking of the similarities measures. The implication of this change is that there is no intrinsic or extrinsic metric associated with the similarities. The change, or transformation procedure is shown in figure V.4. Also, as a simple ordering, 1,2,3,...,n(n-1)/2, the $\hat{d}_{ij}$'s are simpler to work with.

Figure V.4

Kruskal's transformation procedure for the raw similarities.

$$\{\delta_{ij}\} \xrightarrow{\text{monotonic}} \{\hat{d}_{ij}\} \xleftrightarrow{\text{monotonic}} \{d_{ij}\}$$

similarities — transformed similarities — distances

Kruskal also desired a more mathematical explanation. He likened the problem to statistical curve fitting. Consequently, he proposed, (in place of Shepard's S statistic above, in which one ends up multiplying similarities (Kruskal, 1964a, p. 8)), a statistic to evaluate and guide the iterations, termed STRESS, and denoted, S, where,

$$S = \left(\frac{\sum_i \sum_j (d_{ij} - \hat{d}_{ij})^2}{\sum_i \sum_j d_{ij}^2} \right)^{1/2}$$

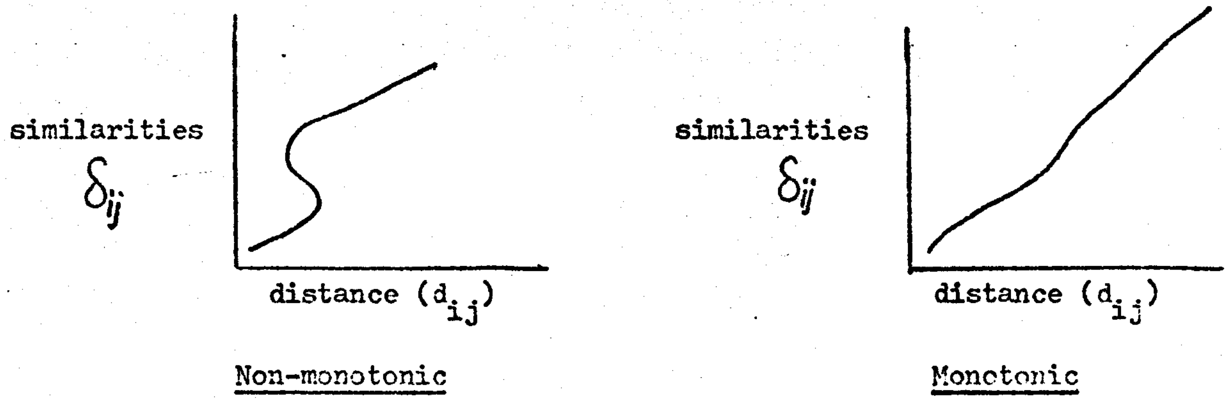
Clearly, this term is similar to the root mean square and follows the least squares approach of statistical curve fitting. It has several important properties, but first, let us digress somewhat.

Consider a graph (Figure V.5) with distance along one axis and similarities along the other. We want a configuration in which similar stimuli are close together. Therefore, we would like a curve rising from the origin to the right, or in short, a perfect monotone fit. That is, no point would be above and to the left of another point. All would be to the right or above as in figure V.5.

Now, starting with an arbitrary configuration, we would expect some violation of this monotone criterion, as in figure V.6, where the order of distances to A, B, and C, do not match the order of similarities. But, if

Figure V.5

Monotone and Non-monotone relations.

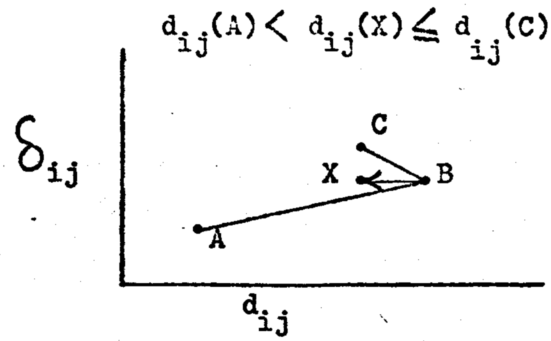


Note: Shepard uses similarities, δ_{ij}; Kruskal uses the transformed similarities, $\hat{d}_{ij}$.

we shift point B to X, then our perfect (weak) monotonic relationship is restored.

Figure V.6

The iterative procedure for calculating the new interpoint distances.



Note: Each point (eg. A, B, C) represents the "distance" between two stimuli, i, j.

Returning to the stress value, S, the distance associated with the original location of B, is defined as d_{ij}; the new location, $\hat{d}_{ij}$. The numerator, then, is a measure of how badly the monotone requirement is violated. As $|d_{ij} - \hat{d}_{ij}|$ increases, the value of S increases, and the goodness of fit declines. As $\hat{d}_{ij}$ approaches d_{ij}, goodness of fit improves. This statistic then, also known as a loss function, provides a good measure of fit in respect to the monotonicity requirement.

A second property is that this statistic is invariant under stretching or shrinking of the distance measure. The denominator, the d_{ij}^2 term, normalizes the statistic so that it is independent of the unit of measure, or any (common) transformations.

Finally, the statistic has a third desirable property. While it is invariant with respect to the distance measure, and with respect to any monotonic transformation of a given similarity matrix, it does vary continously with changes in the coordinates of the configuration (Shepard, 1966, p. 293). This permits the calculation of partial derivatives so that each point is represented by an r-tuple (r=dimensionality) of partial derivatives. This r-tuple is, in fact, the gradient at that point, which, when equated to zero, determines the direction in which the point should move to increase monotonicity.[1] In this way, Kruskal was able to provide a much more elegant algorithm for solving the multidimensional scaling problem. This approach has become known as the method of steepest descent.

The mathematical formulation of the iterative procedure allowed Kruskal to avoid another problem, that of ending up with a solution in n-1 dimensions. The dimensionality, 'r', is stipulated before calculations begin. A best solution is derived in the 'r' dimensions and a stress value determined (a final stress value occurs when the change in S is below some cutoff value; see Figure V.3). The procedure is then repeated with a new 'r' value. The 'r' value is varied to obtain as many solutions as seem desirable.

The problem then arises of determining the "proper" dimensionality. One would like that solution with the smallest stress value and the smallest dimensionality. Clearly, if the dimensionality is n-1, the stress value is zero, but you have learned nothing new. The object is to derive a configuration which parsimoniously represents the data so that interpretations are possible.

This problem is a serious one, and not at all resolved. Kruskal suggests three "aids" for determining the dimensionality:

1) Graph the stress values against the number of dimensions. An elbow in the curve indicates that the addition of one more dimension does not contribute greatly to reducing stress. Note that stress is inversely related to the number of dimensions.
2) Select that dimensionality which permits interpretation. If more dimensions do not greatly reduce stress and are uninterpretable, they may be ignored.
3) If there is an independent estimate of the statistical error in the data, compare it to the stress value.

These criteria can be criticized simply on their subjective nature. However, they do provide a rule of thumb. And, along with the absolute size of the stress values, provide an indication of how many attributes are being used by individuals to make judgements on original stimuli.

---

1. For a clear description of the concepts of, partial derivatives, directional derivatives, gradients, and their application in maximization and minimization problems, see Flanigan and Kazdan, 1971, Chapter 5.

Since the Kruskal-Shepard developments, numerous other contributions have been made, although mostly of a technical nature. The conceptual framework has changed very little.

Guttman (1968), working independently of Kruskal, has developed a nonmetric approach with a much more developed mathematical formulation. In short, his approach involves a two phase iterative process where first the $\hat{d}_{ij}$'s, the transformed similarities (see above), are fixed and the d_{ij}'s, the distances, are manipulated, and then the d_{ij}'s fixed, and the $\hat{d}_{ij}$'s manipulated. Kruskal utilized a single-phase procedure. The advantage of the Guttman approach is that it ensures convergence to a minimum. Kruskal's approach does not guarantee (mathematically) convergence, but in practice is highly successful. The Guttman approach, also, utilizes a non-arbitrary initial configuration which speeds up the procedure. Kruskal utilized an arbitrary or pseudo-random initial configuration which can increase computer time as well as increase the possibility of converging to a local (not absolute) minimum.

The work of Guttman has been computerized by James Lingoes, and hence this approach, formally named Smallest Space Analysis, is often referred to as the Guttman-Lingoes approach. A technical description and comparison of the Kruskal and Guttman-Lingoes algorithms is provided by Lingoes and Roskam (1970).

Other algorithms also exist. These include those by McGee (1966; 1968) and Young and Torgerson (See, for example, Young, 1968), but these are similar in structure to the two main approaches.

Before proceeding to an examination of certain problems in MDS, we might briefly summarize the nonmetric approach. Given n stimuli and n(n-1)/2 similarities measures (missing data can be accomodated, but we consider only complete data matrices), we conceptualize, 1) the n stimuli as points in an assumed dimensionally organized metric space, and 2) the n(n-1)/2 similarities measures as indications of the distances between the n points, such that similar points are close together. An initial configuration is derived, either arbitrarily or through preliminary calculations, interpoint distances are calculated, and the order of distances are compared to the order of similarities. The points are then shifted so as to reduce the stress, that is, to improve the monotonic relationship; new distances are calculated and monotonically regressed against the similarities again. The iterative procedure continues until stress is no longer reducible, given the number of dimensions. The process is repeated for several dimensionalities. In each case the result is a configuration of points in space, so that, given r dimensions, each stimulus (point) is characterized by an r-tuple of numbers, or, the r coordinate values.

We now may ask, what use is this configuration? Immediately there would seem to be two possibilities, 1) the scale values can be used as input to some exogenous technique or theory, or 2) inferences can be made directly from the derived structure.

V.5 MDS: Technique or Criterion?

Shepard (1962) originally considered MDS as simply a method of data reduction, that is, a means of economically representing the structure in the data. In effect, this represents a type of technique (like factor analysis), whereby the objective is to obtain a meaningful scale for use in a further technique or theory. The scale itself is not interpreted.

Shepard quickly realized, however, that one could make inferences directly. The pattern of points could be interpreted as lying in a dimensional space from which attributes could be inferred. In this regard, MDS operates as a criterion. Its internal structure operates as a model of perception; hypotheses can be formulated and tested.

Looked at another way, MDS used as a technique does not question the data. The objective is merely to provide a scale. Hence, it is insensitive to error. MDS used as a criterion, questions the data, that is, do the data match the theory inherent in the MDS approach. Consequently, it is sensitive to error (Coombs, Dawes, & Tversky, 1970).

As an example, a researcher may want to test the hypothesis, derived in some theoretical formulation, that individuals evaluate grocery stores only on the bases of price and service. Using similarities data for several stores, he could then evaluate the validity of his hypothesis. A low stress value in two dimensions in which stores seemed to be arranged along the dimensions of price and service would support his contention. In this instance, the researcher is assuming the inherent model of perception of MDS. Consequently, MDS is operating as a criterion. It is this use of MDS which we wish to explore more fully.

V.6 Problems of Applying MDS as a Criterion

Basically, there would seem to be four major conceptual problems in using MDS as a model of perception. These are:

1) the problem of dimensionality, that is, determining the minimum, appropriate, number of dimensions of the space;
2) the problem of interpretability, that is, how the axes are to be located and interpreted;
3) the problem of homogeneity, that is, how uniform are individuals' perceptions;
4) the problem of intransitivities, that is, are we justified in representing psychological concepts without intransitivities?

The problem of dimensionality is discussed above and will not be further elaborated, except to point out that if, as in the example above, the exact number of dimensions is conjectured a priori, the problem may not exist; only one specific dimensionality need be examined. However, since few theories exist (let alone include specific statements about dimensionality), most research is largely exploratory so that the problem is a common one.

The problem of interpretability is conceptually more serious. Having derived a metric space, it is then necessary to locate an origin and principal axes. If Euclidean space is assumed, interpoint distances are invariant under all rigid transformations (i.e., translations, reflections, and scalar multiplications), so that selection of an origin and axes are purely arbitrary (i.e., dimensions are unique up to an interval scale; interpoint distances are unique up to a ratio scale). Therefore, the axes can be rotated to any position to facilitate interpretation. However, if we relax the assumption of Euclidean space, permitting any Minkowski space (see IV.2), rotation is not permitted. Distances will not be invariant under rotation.

Essentially, we must realize that the selection of a specific metric implies a specific combination of the attributes. For example, the Euclidean metric implies that attributes are added proportionally, whereas the city-block metric implies unweighted addition of the attributes. The lack of theory, along with the advantage of rotation and familiarity, has resulted in almost exclusive use of the Euclidean metric. But, as we pointed out earlier (IV.2), there is no psychological support for selecting the Euclidean metre. This leads one to ponder, 'how might the results change if we vary the metric space?'. Some preliminary work has been done (Cross, 1965; Demko and Briggs, 1971), but the conceptual problem has not been resolved and remains an important research question.

Closely connected with the problem of rotation is the problem of labelling. Assuming Euclidean space, let us reexamine the grocery store example above. We would like to rotate the axes so that one dimension would represent variation in service, the other dimension would represent variation in price. In general we would like to rotate the axes to an interpretable position, that is, so that the dimensions would represent meaningful attributes of the phenomena. To make this type of interpretation, exogenous information is required. For example, specific information is required concerning price and service for each store. Then we could construct the axes so that the stores with high prices would project close to each other on one dimension and stores with low prices would project close to each other at the opposite end of the dimension (assume a central position for the origin). The second dimension would be necessarily perpendicular. If this order of projections could not be obtained, we would be forced to conclude that price was not a relevant attribute.

The important point is that to interpret any spatial configuration we need to be able to compare the order of the projections of the stimuli on dimensions with some exogenous information. Of course, there is always the possibility that the exogenous information we choose to collect may correlate well with the order of projections, but not be the relevant attribute. While this is important, I do not want to overstate the point. Most researchers should be reasonably familiar with their phenomena, qualitatively at least, to avoid this problem. But with some exploratory work, especially with complex phenomena, spurious interpretation is possible.

One method of preventing this problem is evident in the choice model proposed by Briggs and Demko (1970; 1971). As well as collecting similarity data, one can also collect ranked data on a variety of specific attributes. These are then collapsed into factor scores which are compared to the coordinate values obtained in the MDS via a canonical correlation. One then has specific information about each stimulus, and about other stimuli that it is grouped with, as well as another statistical measure of the goodness of fit of the MDS configuration.

The problem of interpretability, then, may be divided into two questions, 1) can the axes be rotated?, and 2) what exogenous information is required? Obviously, MDS is most useful when testing a specific theory which directs the selection of the exogenous information.

Recently, another method has been suggested which reduces the importance of the dimensions in interpreting the configuration. One attempts merely to label regions. Johnson (1967) has developed a clustering technique based on interpoint distances. The important feature is that distances remain meaningful after clustering. That is, not only are the distances between points defined, but after clustering, the distances from the cluster to other points (or clusters) are also defined. In this way cluster regions can be defined. Further, he provides two techniques; one for maximizing compactness, and one for maximizing connectedness. It is anticipated that the use of clustering will become more widespread in promoting dimensionless interpretations.

The problem of homogeneity is related to how people perceive phenomena. MDS explicitly assumes perceptions are homogeneous. Individuals perceive a set of (comparable) objects along the same set of dimensions. Some individuals may have extra idiosyncratic dimensions which are relevant to them alone, but in general, it is assumed that individuals perceive objects in "pretty much the same way". Notice also, in deriving a group stimulus space, not only is there the assumption that individuals are using the same dimensions, but, that each phenomenon is ordered or weighted along each dimension similarly by everyone. That is, each dimension has the same relevance or salience to each person, so that the group space, obtained from some averaging of individuals is assumed relevant to each individual.

These are strong assumptions. Recently, Green and Carmone (1971) have documented numerous violations of these assumptions. They suggest, for example, that perceptions may depend on personality and cultural variables, as well as the context or outside conditions present (i.e., scenario dependent) in eliciting the data.

To avoid violating these assumptions, one needs to concentrate on groups with similar perceptions. One method (Demko, 1971) is to factor the individual similarities arrays to obtain more homogenous subgroups upon which to perform the analysis.[1]

As well as the problem of inter-individual comparability, the problem of homogeneity includes the problem of within individual, inter-phenomena comparability. That is, does an individual use the same dimensions for each phenomenon. For example, an individual may see a red apple as being similar to a red cadillac. Also, he may see a red apple as being similar to an orange. But, he is not likely to see an orange as being similar to a red cadillac (Stefflre, 1971). This example is admittedly an extreme one, but it emphasizes that we need comparable phenomena, that is, various types of apples, or various

1. INDSCAL also offers a method of obtaining groups with homogenous perceptions. See section VII.

types of fruits, or various types of cars, but not all three. The problem, of course, is that we are never quite sure whether other people see two objects as being in the same set the way we do.

This example also emphasizes the problem of intransitivities.[1] Letting A, B, C, be stimuli and R be the relation, "is similar to", then ARB and ARC implies BRC. This is transitivity. Stefflre (1971) believes, that in fact, psychologically there are probably numerous intransitivities even among seemingly comparable items. But, the assumption of a metric space precludes intransitivities. Letting A be the red apple, B the red cadillac, and C the orange, we have the example above. But, ARB and ARC does not result in BRC, as it should logically. This intransitivity violates the triangle inequality, a corner stone of metric spaces. Similarly, Stefflre points out that often the other metric properties, positivity and symmetry, also fail. Confusion data, for example, where the number of times two items are confused are recorded as similarities data, may indicate that indeed, $ARA \neq 0$. 'A' may not be close to itself. 'A' may not be confused with B to the same extent B is confused with A. Then, $ARB \neq BRA$, which violates the property of symmetry.

Stefflre's arguments have a strong intuitive appeal. However, since we do not know how the mind works exactly, as we might, say, a car engine, we cannot prove or disprove his arguments, nor our model, purely on logical a priori grounds. We are never sure of "the unique truth". His cautions, however, cannot be taken lightly. Perhaps he is really saying that we are dealing, in part, with dialectic concepts (see Georgescu-Roegen, 1971); that is, concepts which can be A, not-A, and both A and not-A, at the same time. This violates the principle of contradiction. An example is the concept 'democracy'. A nation can be said to be 'democratic', or 'non-democratic', or to have features which are both 'democratic and non-democratic'. The conclusion, therefore, is that we must be very careful to choose comparable objects which are distinct identities, and use similarity measures which least violate the metric assumptions. We cannot hope for perfection, merely reasonably supported findings.

---

1. Intransitivity is the case, ARB, ARC, but, BnotRC. This may be seen clearly with order relations, for example, $A > B$, $B > C$, but $C > A$.

VI Multidimensional Unfolding

VI.1 The Problem

Given n stimuli and the ranked choices of a number of individuals on those n stimuli, determine a set of numbers for each stimulus, and for each individual, so that all relationships (rankings) between individuals and stimuli are represented. Both stimuli and individuals are conceived of as points in a common metric space --- hence the name joint space --- of r dimensions, where r represents the number of relevant latent attributes inherent in the stimuli, and where the distance between any individual point and stimulus point indicates the degree of preference which that individual has for that stimulus.

In effect, the problem is to derive a joint ordering of individuals and stimuli, given only each individual's preference rankings. Each individual's point is considered his ideal point --- that combination of attributes which he would most prefer --- with preference declining away from the ideal, monotonically, and symmetrically (assuming Euclidean space). Each person's utility (preference) function is single peaked at his ideal point.

VI.2 The Concept of "Unfolding"

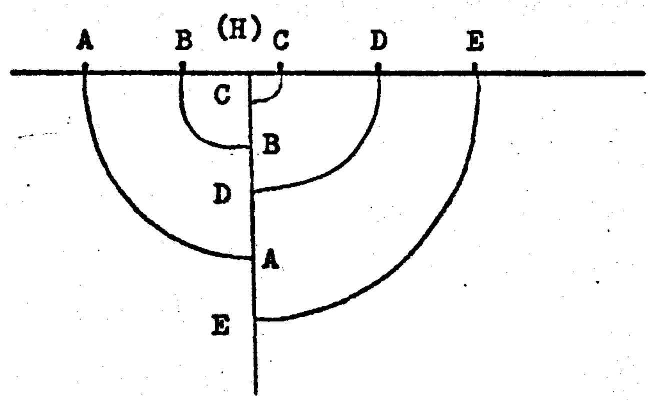
Coombs (1964) first coined the term "unfolding". The concept is related to the notion of an ideal point.

Consider the unidimensional case where the objective is to locate several individuals on one common dimension. Each individual's ranking of stimuli, say, ABCDE, is known as his I scale. The common dimension is the J scale (Coombs, 1964). The I scale corresponds to the J scale folded at the ideal point. For example, given the J scale ABCDE, and the I scale, CBDAE, we could obtain the I scale from the J scale by folding the J scale on the C side of the midpoint between B and C, as in figure VI.1. Since the data are in the form of I scales, the problem is to unfold the I scales to obtain the J scale.

Now it is easy to see that one J scale can accommodate many preference rankings, both by varying the ideal points and the inter-stimulus distances. However, the above unidimensional example cannot accomodate the ranking, for instance, CDEBA. Then, one can assume that the preference ranking is a function of more than one attribute, so that a multidimensional solution is

Figure VI.1

Unfolding the I scale, CBDAE, to the J scale, ABCDE, with ideal point (H).



required. To conceptualize this unfolding, Coombs (1964, p. 141) suggests that the I scale can be derived by,

> picking up the plane as if it were a handkerchief at the ideal point and compressing the handkerchief into a line.

Coombs' unidimensional unfolding model has been extended to the multidimensional case by Bennett and Hays (1960). However, their method is only able to derive a rank ordering of the stimuli and ideal points on each dimension. More recently, the algorithms of Kruskal (1964), Young and Torgerson (see Young, 1968), Carroll and Chang (see Carroll, 1971; Coxon, 1972), have been extended to apply to preference data. These algorithms yield metric information. The properties of these nonmetric unfolding approaches are the same as those for similarities analyses. Namely,

1) Given the preference rankings of a set of stimuli, in which preferences are interpreted as distances from an ideal point, the projections of the stimuli and ideal points can be determined on an arbitrary set of orthogonal axes.
2) A metric space can be recovered from the nonmetric information.
3) The choice of judgment attributes is unspecified, although the context, preference, is specified.

VI.3 Assumptions and Criticisms

The basic assumptions of the MDU model coincide closely with the MDS model. The MDU assumptions are listed below.

1) Individuals <u>perceive</u> stimuli as some combination of attributes. The <u>weighted</u> combination of attributes determines preference (i.e., the weights refer to preference weights).
2) Stimuli and individuals can be represented in a common, joint, metric space.

3) The space is dimensionally organized so that the orthogonal dimensions represent the weighted, latent, attributes of the stimuli.

4) Stimuli and individuals are located on the scale according to the effects of the weighted attributes.

5) The distances in psychological space between ideal points and stimulus points measure the preferences for each stimulus, for each individual.

6) The resulting configuration is valid for all individuals. That is, all individuals perceive, and weight, the stimuli similarly.

7) All stimuli are comparable.

The basic distinction between the MDU conceptualization and the MDS conceptualization, besides the idea of a joint versus a stimulus space, lies in the treatment of the attribute dimensions. MDS attempts to recover the attribute space without reference to value. An object may possess both positive and negative attributes. This should not affect perception. MDU, on the other hand, explicitly attempts to incorporate value judgments.

More precisely, the stimuli are conceived of as a union of attributes (as in MDS), where each attribute is weighted by the specific value attached to that attribute by all individuals. Hence, the relative importance of some perceived attributes may dominate the preference ranking. Other attributes may have little or no value attachment.

For example, a city may be perceived as being made up of two dimensions --- a social dimension and an economic dimension (Demko, 1971). Individuals, however, perhaps with a low income, while recognizing the social dimension, will value it lowly, and at the same time, value the economic dimension highly. Hence, a unit distance along the economic dimension in perceptual space will be increased in preference space, while a unit distance along the social dimension in perceptual space will be reduced upon transformation into preference space.[1] Thus, preferences will tend to reflect the economic dimension; economically buoyant cities will rank high, while economically depressed cities will rank low. The social dimension may be used only to differentiate between equally buoyant or equally depressed cities.

More generally, Green and Carmone (1968) have examined the relation between MDS and MDU using experimental data. Specifically, their question was: 'how does the configuration of stimuli derived using similarities data differ from the configuration of stimuli derived using preferential choice data?'. They conclude:

---

1. The weights alter the perceptual dimensions so that a unit distance on one dimension is preferentially equivalent to the same distance on all other dimensions in the preference (joint) space.

> In effect, what the unfolding approaches (Torgerson and Young; Kruskal; Carroll and Chang) appear to be providing is not the standard configuration (obtained from similarities data) but a transformed configuration in which the interpoint distances between stimulus pairs are weighted by the "importance" attached to each dimension in the preference context. (P. 29)

The dimensions are the same, they have just been stretched or compressed (and possibly rotated) depending on the importance. Notice, however, the dimensions are group-weighted so that individual differences in preference are accounted for by varying the ideal point, not the weights of the dimensions. However, the stimulus configuration is assumed representative for each individual in the group.

The MDU model, then, allows for individual differences in preference, but does not allow for individual differences in perception (assumption 6), or preference saliences (i.e., different preference weights). These parallel the crucial assumptions of MDS, although for slightly different reasons. With similarities data, using MDS, the configuration is derived from a single combined similarities matrix. Therefore, there are no individual similarities in the actual analysis, and one must assume homogeneity. With preferential choice data, using MDU, however, the procedure maintains the individual data and includes as output, ideal points. Hence, we can see the relationship of the individual to the entire configuration. But, if both preferences and perceptions are allowed to vary, the distances become difficult to interpret. Variations could be a result of either differences in preference or differences in perception.[1] Or as Green and Carmone (1968) comment succinctly:

> ... all unfolding techniques assume that respondents' perceptions are homogeneous ... otherwise differences in perception would be confounded with differences in preference. (P. 18)

They further note that homogeneity is required since it is assumed that the variables in each utility function (the arguments) are the dimensions of the space, that is, the attributes common to all individuals.

Expanding on this notion of utility functions implicit in unfolding techniques, Green and Carmone (1968) point out that various algorithms make different assumptions. They emphasize first, that the MDU model derives a configuration of stimuli and individuals, which includes stimulus-stimulus and person-person distances without any information constraints in the original data. Hence, the question arises, given N individuals and n

---

1. A similar argument relates to 'preference saliences' in the present case, which is a form of internal analysis. This limitation may be removed in some preference models if they use an external analysis. See section VI.4.

stimuli providing an Nxn matrix, are within column comparisons meaningful, that is, are inter-stimulus comparisons reasonable?

Torgerson and Young, by assuming that all cell values are comparable, both within rows and within columns, are implicitly assuming that all individuals have the same symmetric, monotone decreasing utility function. On the other hand, Kruskal uses only within row comparisons, thereby permitting different utility functions, although of the same form, for different people. The monotonicity constraint is satisfied only "within people". This reduces the number of constraints in locating the configuration, but does not imply any interpersonal utility comparisons as does Young and Torgerson's method.

Carroll and Chang provide a method with even fewer constraints. It permits multipeaked, asymmetrical utility functions. However, this method does not provide a joint space configuration. More will be said about this model below.

The conceptualization of the utility function in the MDU model is perhaps one of its most interesting aspects. Coombs originally conceptualized the function as single peaked, symmetric, monotonically declining throughout and approaching zero at an infinite distance. The normal utility function, in contrast, commences at zero and rises throughout. Consequently, Coombs labelled his function a "disutility" function.

This disutility function can be turned into a utility function again by considering each individual point an anti-ideal point. This is only relevant, of course, for some attributes. Carroll (1971) suggests that, in fact, the attribute space might include both positive and negative attributes. An individual point would indicate the ideal for some dimensions, and the anti-ideal for others. This would form a saddle point. The resulting utility function would be asymmetric.

Also, there is the possibility of multi-peaked functions. People like either hot or cold tea, but not luke-warm tea (Carroll, 1971). This can be represented by a simple anti-ideal point. However, people probably do not like tea either too hot or too cold. This would require a two-peaked (bimodal) function. This cannot be handled by any method at the moment.

The unfolding model, then, provides possibilities for examining utility functions, at least in some of the simpler cases, with the use of simple ranked data. In particular, it focuses on the perceived attributes of a number of related objects, rather than the objects themselves, as is more conventional in economics. This is an important advantage. There is one difficulty, however. The utility information generated is ratio scaled, which gives the utility function a "cardinality" quality. In effect, this approach provides a cardinal measure of utility, implying that all combinations of attributes are comparable. This can lead to considerations of irrelevant alternatives (see Arrow, 1967), which does not fit well with decision theory postulates (see, for example, Isard, 1969).

VI.4 Related Individual Preference Models

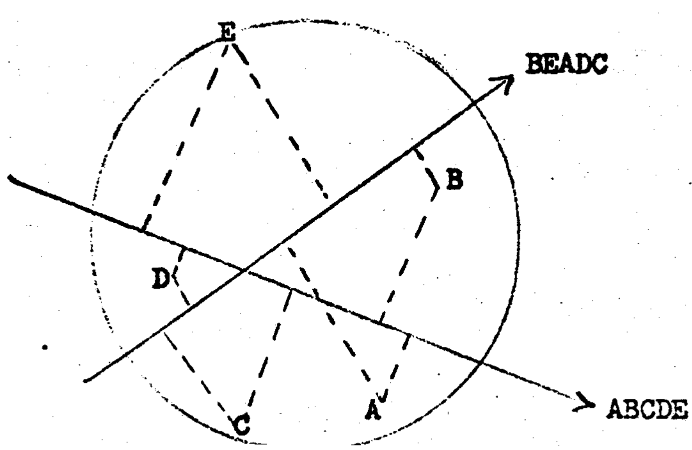
As well as Coombs' "unfolding" model, at least three similar models are concerned with individual preferences; Tucker and Messick's "point of view" or "vector" model, Carroll and Chang's "general unfolding" model, and Carroll and Chang's "weighted unfolding" model. Both the general and weighted unfolding models are included in Carroll and Chang's computer program, PREFMAP, as Phase I and Phase II, respectively.

These models differ from the strict unfolding model in that they involve an external analysis. Unfolding uses simply the individual preference data to determine the final configuration without reference to any outside stimulus dimensions. Hence, it is internal analysis. The vector model and PREFMAP use an a priori set of dimensions, usually determined from a preliminary MDS analysis of similarities data. Thus, the stimulus dimensions are determined externally, and then individuals related to that space by their preferences (Carroll, 1971, p. 249).

The vector model conceptualizes the individual as a vector, or directed line segment through the multidimensional stimulus space. A subject's preferences are reflected by the ordering of stimuli projected onto his vector. An example is shown in figure VI.2.

Figure VI.2

Tucker and Messick's Vector Model, showing the 'vectors' for two individuals; the order of the projections of the stimuli represents the order of preferences. Source: Carroll, 1971, p. 250.



The relative importance that each individual attaches to the dimensions of the space (determined externally) can now be determined. The cosines of the angles between the individual's vector and the dimensions of the space, indicate his personal weighting. This could not be determined in the Coombs model, since we were dealing with a stimulus space already weighted.

However, the two models are actually quite similar.[1] The vector model assumes preferences increase with all dimensions,(assuming the ideal point has been translated to the origin). In effect the vector model simply assumes the ideal point is infinitely far away. Conceptually, the two models are almost mirror images.

The general unfolding model, PREFMAP, Phase I, attempts to resolve two problems in the strict unfolding model. Both problems are related to homogeneity. One problem is the assumption that a given difference on a dimension means as much to one individual as to another. The second problem is that all individuals use the same set of preference dimensions (Carroll, 1971). In effect, the first problem is related to the units of a given dimension --- does $10 mean the same to all people --- while the second problem is related to the location of the dimensions, that is, rotation.

The general unfolding model, then, assumes a common perceptual space but permits each individual to choose a unique set of reference axes in the preference space, that is, to rotate the axes of the space, and to weight the dimensions (i.e., to stretch or compress the configuration). The immediate effect of weighting on the configuration is to distort the disutility function. Instead of circular isopreference lines, they become ellipsoids (or hyperellipsoids). A large weight on a dimension for an individual (i.e., that dimension is very important) compresses the configuration along that axis. A small weight stretches the configuration. With weighting and rotation, there is no common joint space. With n individuals, there are n configurations. Without rotation, but with weights, a common space can be determined, although isopreference lines for each individual must be shown. Without individual weights, with or without rotation, the general model reverts to the Coombsian unfolding model.

The weighted unfolding model, PREFMAP, Phase II, is simply the unrotated general unfolding model. The Phase II model permits individual weightings but not individual rotation. Each individual is assumed to use the same preference axes to evaluate the stimuli.

---

1. Carroll suggests the vector model is a special case of the unfolding model. However, the unfolding model uses a different stimulus space.

At this point it is important to clarify several distinctions concerning preferences, perceptions, and weights or saliences. All the preference models assume that individuals have common perceptions. That is, a group stimulus space is valid for all individuals. More specifically, all individuals perceive the stimuli alike, using the same dimensions with the same perceptual saliences on those dimensions. More will be said about perceptual saliences in section VIII. The preference models differ, however, in their treatment of preference dimensions and weights. Simple unfolding assumes common dimensions and weights for all individuals. The perceptual dimensions are distorted by a group weight.[1] The weighted unfolding model, PREFMAP, Phase II, assumes common dimensions, but permits individual weightings of the dimensions. It is easy to conceptualize these weights as distorting the perceptual dimensions, so that if an individual values one dimension more than another he will assign the more valuable dimension a larger weight. This tends to compress that dimension so that stimuli projecting highly on that dimension are pulled closer to the individual's ideal point. The general unfolding model, PREFMAP, Phase I, permits each individual to rotate his preference axes before weighting them. These "private dimensions" may be a combination of original perceptual dimensions which the individual places a high value on. In this way, for example, a two dimensional perceptual space may be converted into a one dimensional preference space. The new single dimension would represent a weighted (evaluative weight) combination of the original perceptual dimensions (see Coxon, 1972, p. 12).

Consequently, the general, weighted, and simple unfolding models offer different advantages. The general model permits greater focus on the individual and would be useful in personality, social, and economic studies, in which the object is to relate two sets of data about the individual. On the other hand, the simple unfolding model permits greater focus on aspects of the stimuli and would be useful in studies directly interested in the aspects of objects which are, in general, liked and disliked. The point is, both the general and the simple unfolding models are useful depending on the objectives. The weighted unfolding model offers an option of intermediate complexity. It is more flexible in its assumptions than simple unfolding, less flexible than general unfolding, and correspondingly intermediate in interpretability.

---

1. Recall that in the simple unfolding case, the dimensions are determined internally. Hence, the procedure does not involve first calculating the perceptual dimensions and then weighting them, but the dimensions are calculated in one procedure. The distortion from the perceptual space can [illegible]

VII Individual Differences Scaling (INDSCAL)

The previous two sections (V, VI) were devoted to multidimensional techniques related to the scaling of perceptions and preferences respectively. MDS is designed to provide the perceptual configuration of a number of comparable stimuli. However, a basic assumption is that individuals perceive a unit distance on one dimension the same as a unit distance on another dimension, or simply that individual perceptions are homogeneous. INDSCAL is a technique which attempts to relax this assumption, namely, that all individuals attach the same perceptual weights to all dimensions, although it maintains the assumption that all individuals use the same set of dimensions. That is, unlike the "general unfolding" preference model counterpart, INDSCAL does not permit individual rotations. It does, however, provide a measure of how well the configuration accounts for each individual's perceptions so that individuals with extreme "points of view" may be identified and analyzed separately.

VII.1 The Problem

Given n stimuli and the ranked similarities measures on these stimuli by a number of individuals, determine a set of numbers for each stimulus, and for each individual, so that, all relationships (rankings) between individuals and stimuli are represented. The stimuli are conceived as points in a metric space of r dimensions where each dimension represents a perceived attribute inherent in the stimuli. Individuals are conceived as points in a subject space of the same dimension as the stimulus space, so that each point indicates that individual's "perceptual saliences" or perceptual weights.

The similarity between the problem formulated in INDSCAL and the problem considered in MDS is clear. The distinction rests on the treatment of individual similarities. MDS combines all the individual similarities matrices into a single "average" matrix. Effectively, this assumes that perceptions are homogeneous. INDSCAL on the other hand, maintains the individual matrices in the analysis. The objective, in fact, is to uncover the individual differences in perception.

VII.2 The Model

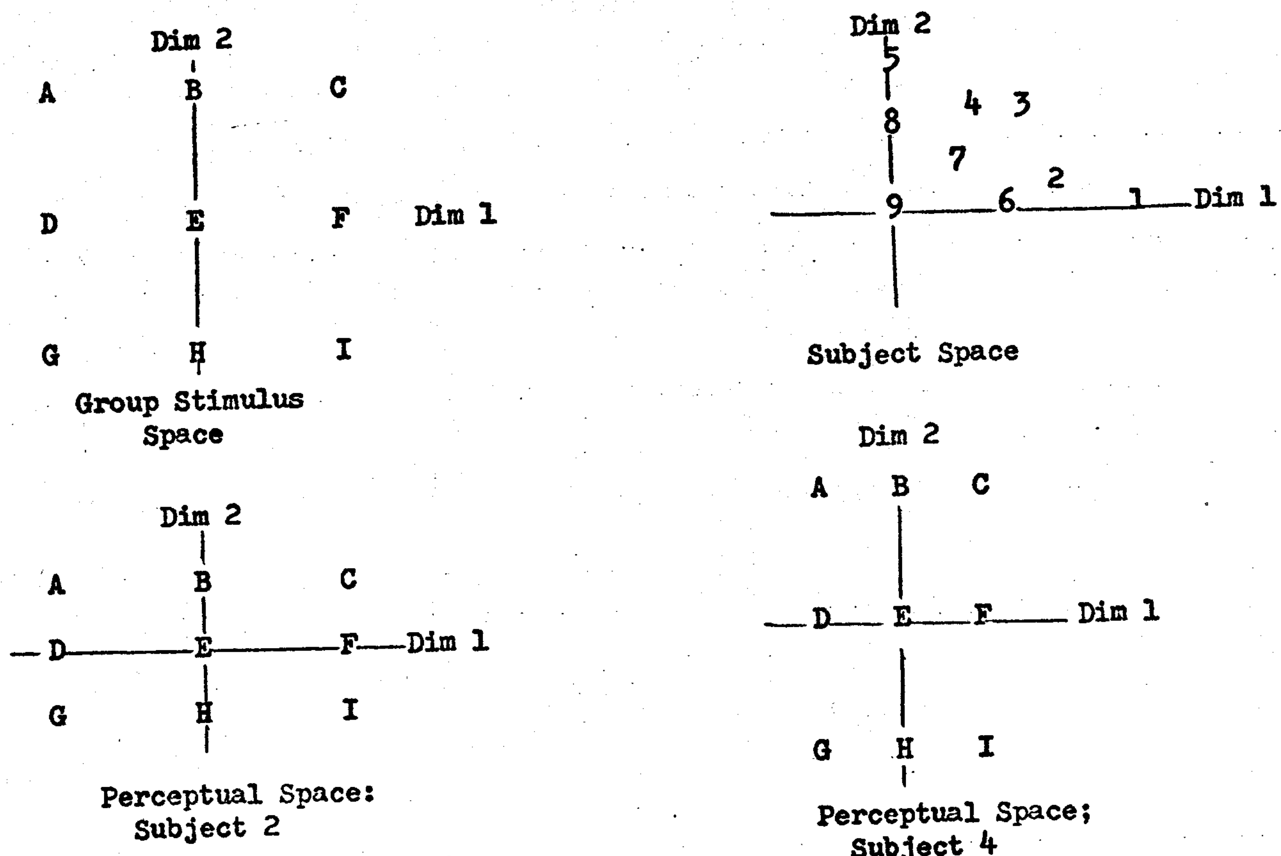
INDSCAL is a technique for scaling individual perceptions which is closely related to MDS. First, it makes all the assumptions of MDS, except for assumption 6 (V.3), concerning homogeneity. Instead, the derived stimulus configuration, the group stimulus space, is assumed not valid for all individuals. Consequently, INDSCAL derives a related subject space to represent the individuals. Second, INDSCAL maintains the properties of MDS except for one deviation. INDSCAL has two versions, a metric and a quasi-nonmetric. The metric approach assumes that distances are some linear function of the similarities. The quasi-nonmetric assumes, like MDS, that distances are some monotonic function of the similarities. The metric version seems to be more widespread in use despite the more stringent requirement concerning the data. One can proceed, however, by performing an MDS analysis for each individual, thereby obtaining ratio scaled distances as input to the metric routine. In this way, INDSCAL becomes quasi-nonmetric. There are costs in computer time, however.

In the subject space, individuals appear as points. The coordinates of these points indicate the relative weight, importance, or "perceptual salience", of each dimension in the stimulus space. Consequently, the number of dimensions of the subject space correspond to the number of dimensions in the stimulus space. In effect, these weights represent a distortion factor. They indicate how each dimension of the stimulus space must be distorted, compressed, or stretched, so that the stimulus space will represent that individual's perception of the stimuli. The process can be pictured as multiplication of the stimulus space by a constant vector representing the individual's saliences for attributes of the stimulus space to produce a personal stimulus space. Figure VII.1 gives an example of a group stimulus space, a subject space, and two personal spaces.

Examining figure VII.1, respondent 2, for example, weighs dimension one more highly than dimension two, so that his personal space is compressed along dimension two. Respondent 4, in contrast, reverses the saliences of the dimensions, compressing dimension one. Respondents 5, 8, 6, and 1, perceive only one dimension, either dimension two or dimension one. Their "maps" could be represented by a straight line. Respondents 3 and 7 represent an interesting situation. Both have proportionally equal weights for each dimension, so that their maps would have identical patterns, both with the

Figure VII.1

INDSCAL Output: Group and Subject Spaces.



Source: Carroll, 1971, p, 243.

group stimulus space and with each other. This degree of similarity between two individual patterns can be thought of as their communality. Two individuals on the same ray fom the origin would have the same pattern and hence a high degree of communality. However, the distance from the origin in the subject space is a measure of the variance or explanation accounted for by the analysis (actually the measure is the square root of distance). Individuals close to the origin have a greater portion of variance unexplained. Hence, in the case of subjects 7 and 3, their perceptions are similar, except that less of the data for subject 7 have been accounted for. Subject 9, for example, is completely unaccounted for in this analysis.

Carroll (1971) suggests that this unexplained factor may result from two sources, 1) individuals may possess idiosyncratic dimensions outside the analysis and not recorded in the group space, and 2) individual responses may be unreliable.

Finally, INDSCAL incorporates one more feature which is both advantagous and limiting. INDSCAL solves the group stimulus space for a unique orientation of the axes, maximizing the goodness of fit and promoting interpretation. Since individuals weigh the axes differently, each wishes to compress the space differently. Hence, a different orientation of the axes results in different weightings. More precisely, interpoint distances are, now, not invariant under rotation. Therefore, we can select that orientation which maximizes the explanation in the subject space. For example, if one orientation resulted in numerous individuals close to the origin, while another located most individuals away from the origin (in the subject space), the second configuration would be utilized. Generally, Carroll (p. 246) states:

> The family of transformations of the stimulus space generated by differential weights varies as a function of the orientation of axes in the group stimulus space.

The difficulty with this approach is that it assumes, implicitly, that rotation is permissable, so that, only Euclidean space can be used. If another metric were permitted, then the variation in explanation would be a function both of the individual saliences and the metric. What then would be the meaning of the unique orientation, psychologically, if the orientation factor, resulting from the metric, dominated the selection of the axes?

To summarize, INDSCAL relaxes the assumption of strict homogeneity of perceptions, although it does suppose a certain degree of communality of perceptions (Wish, 1971). In the process, both a group stimulus space and subject space are defined, so that individual saliences of dimensions can be ascertained and private spaces recovered. Clearly, the emphasis is on the individual, but the unique orientation of the axes which results in the stimulus space promotes interpretation at the group level. In this regard, separating individual and group configurations, INDSCAL may be of considerable use in examining the aggregation problem outlined in section II. In effect, INDSCAL combines two levels of focus, the individual and the small group, and provides comparable results at both levels.

VII.3 Perceptions and Preferences: A Clarification

In section VI we distinguished between the "simple" and "general" unfolding models by reference to preference axes and evaluative weights. Similarly, in this section we have attempted to point out the differences between simple MDS and INDSCAL by reference to perceptual weights or saliences.

We have, in effect, in the course of these discussions conceptualized a very general model of individual decision-making or choice.

We have considered an individual comparing two stimuli. He perceives each stimulus as a combination of attributes which we have considered as dimensions in our model. But we recognize that another individual may not perceive the same two stimuli in exactly the same way. He may not use the same dimensions or attributes as the first individual, or he may attach different perceptual saliences to the dimensions. For example, both individuals may consider that two stimuli, A and B, are very similar. Also, both individuals may be using the same attribute dimensions. However, one individual may consider A similar to B because they possess similar amounts of attribute one, attribute two being irrelevant. While the other individual may feel the stimuli are similar because they possess similar amounts of attribute two, attribute one being irrelevant. In this case the two individuals have different saliences for attributes. This, however, has not affected the degree of similarity between stimuli, A and B. To see how saliences can affect similarities judgments, consider a third stimulus, C, ranking high on dimension one and low on dimension two. One individual would rank C similar to A and B, while the other individual would not, Different perceptual saliences, then, as well as different perceptual dimensions, can affect perception.

Similarly, we have considered preferences to vary between individuals as a result of differences in <u>preference dimensions</u> and/or <u>evaluative weights</u>, assuming perceptions are homogeneous. Preferences have been conceptualized as resulting from individuals attaching personal values to the dimensions of the perceptual space. In effect, different individuals <u>value</u> one unit of an attribute differently. But we have conceptualized a more general model than this. We recognize that the axes in the preference space may be different than in the perceptual space. Individuals may combine perceptual dimensions to form new preference dimensions before attaching evaluative weights. Differences in preference then, assuming homogeneous perceptions, may result either from different rotations or different evaluative weights.

The entire conceptualization results in an individual being able to rank by preference, a number of stimuli. In this way, if the stimuli represent choice possibilities, the individual can reach a decision, selecting the most preferred alternative. However, two individuals need not reach the same choice, as we have seen. This results from,

1) differences in perceptual dimensions,
2) differences in perceptual saliences,
3) differences in preference dimensions, and

4) differences in preference (evaluative) weights.

The relationship between perception and preference should now be clear. The basic distinction is that perceptions are separate from preferences. The basic idea is that there is a distinction.

MDS assumes both homogeneous perceptual dimensions and saliences. No individual variation is permitted. INDSCAL permits individual perceptual saliences but assumes common dimensions. Radical individuals who violate this assumption (close to the origin in the subject space), can be determined and removed, however.

All the preference models assume common perceptual dimensions and weights. But they vary in regard to preference dimension and weights. The "simple" unfolding model assumes homogeneous preference dimensions and evaluative weights. Individual variation is permitted only by varying the ideal point. The "weighted" unfolding model assumes common preference dimensions (i.e., the same dimensions as the perceptual space), but permits individual evaluative weights. Finally, the "general" unfolding model relaxes both assumptions permitting individual rotation of the perceptual axes to create new preference axes, to which individual evaluative weights can be applied.

VIII Geographic Applications

In section II, the study of human geography was organized into an examination of two classes of variables, place attributes and relative location. As well, it was emphasized that individuals' cognition of these variables would influence their behaviour. Hence it is important to measure individual cognition.

In this section, the applications of multidimensional techniques to these problems, the cognition of relative location or distances, and the cognition of place attributes are reviewed. Further, a distinction was made between identifying the relevant cognitive elements, the problem of analysis, and combining these elements with value judgments to determine preferences, the problem of synthesis. It is now clear that analysis is concerned with perception --- 'What do we perceive?'--- and synthesis is concerned with preference --- 'How do we assign values to elements?'. Therefore, this section is organized in the following way. The first part is concerned with the problem of analysis (perception) and the second part with the problem of synthesis (preference). As well, the first section is divided into those applications primarily concerned with location (distances), and those primarily concerned with place attributes. The operational preference models in the second part, do not use this distinction.

It should be emphasized, before proceeding, that the applications reviewed are not exhaustive. There are probably other studies using multi-dimensional techniques which pertain to, or are relevant to, spatial analyses. However, it is hoped those presented will provide a reasonable variety of applications to demonstr ate the usefulness of multidimensional techneques.

VIII.1 Analysis: Distance Applications[1]

Multidimensional scaling appears useful in two distance applications. It can be used as, 1) a type of map transformation using metric data as initial input, or 2) as a type of perceptual mapping using subjective distance estimates as input. In either case, to use the nonmetric routines, the original data must be in the form of ranked (ordinal) data.

A simple but interesting example of a map transformation is that involving the "Roadmap" problem. Greenberg (1969) used an atlas to obtain the road distances between fifteen pairs of cities in the U.S. This amounts to n(n-1)/2 or 105 measurements. These metric measurements were converted to ranked data, interpreted as similarities measures, and used as input to an MDS routine. Interpretation of the output was simplified by the fact that the dimensionality was two, and the north-south, east-west directions known.

The MDS configuration was then compared to the actual arrangement. Discrepancies were small and could be accounted for by the fact that road mileages are not consistently representative of straight-line or shortest-path distances. The MDS configuration, in effect, defined the "true" relative locations of the cities. It was as if the cities were located on a rubber plane which had been stretched in the appropriate directions until all roads were straight.

This example, although simple, demonstrates the wide applicability of MDS techniques. It need not be confined to psychological data, and as such avoids many of the difficulties necessary in psychological formulations. The roadmap problem, for example, avoids the assumptions of homogeneous perceptions, stimulus comparability, or inherent attributes and their additivity (assumptions, 1, 4, 5,

---

1. Much of the discussion in this section is based on the relevant discussion in Rushton and Golledge (1970), and Miedell (1969). Both describe Greenberg's (1969) "Roadmap" problem; the former cite Tobler (1967) and Tobler, et. al.(1970).

6, and 7; section V.3). Also, the assumptions of a metric space, dimensionally organized, need not be considered, since the space is known to be two-dimensional Euclidean.

The simplicity of this example, however, should not cloud its usefulness. Tobler, Miekle, and Detwyler (1970) for instance, used geobotanical distances between islands in the South Pacific (inferred from a diffusion model of plant species) to derive a new map. This transformed map would then represent the true "botanical distances".

We can immediatesly think of numerous other areas where a map of transformed distances might be useful. For example, we might consider telephone rates as similarities measures and derive "teledistances". These new distances might be useful in a gravity model to explain some types of interaction. Similarly, we might use commodity flows or population flows to derive transformed interaction spaces. We might consider intercity airplane times as a measure of similarity to derive time-spaces. Conjecturing that near places require relatively more time for loading and unloading compared to the total length of the intervening trip, we would expect near places to be stretched out in the transformed time-space, and distant places to be pulled together. Railways might like to compare rail-time maps with air-time maps, including the time to and from airports to see if they are competitive. Cost comparisons could also be made.

In general, all types of interaction measures might be used to define new distances for further analysis. As we have seen in the above examples, the resulting map is some transformation of the "real" atlas map. But are these atlas maps real? In fact, they are a two-dimensional representation of the curved earth. Tobler (1967) has demonstrated (although he used metric scaling), that multidimensional scaling can be useful in transforming the spherical properties of the earth into two-dimensional maps. This output becomes the "real" map which enters our atlas. This is an interesting reversal on the question of reality.

MDS has also been used to examine cognitive distance in an urban setting. Golledge, Briggs and Demko (1969) selected several points in the city of Columbus, Ohio, and asked two groups of graduate students to estimate distances for the n(n-1)/2 interpoint distances. These estimates were interpreted as dissimilarity measures and became the input for the MDS routine.

The basic question they were concerned with was: how is the attribute of distance distorted in the cognitive process? They considered familiarity, which is linked to length of residence and activity patterns, the key variable controlling the distortion. To standardize activity spaces, they selected respondents focused at one point, the university campus. To study the impact of length of residence, they selected two groups, one made up of first year graduate students, and one of third year graduate students. Further, they hypothesized that the different lengths of residence would account for intergroup variances, while different activity patterns would account for intra-group variances.

To match orientation of the MDS output with the real world dimensions, points were located largely on two main arteries intersecting at the campus. The derived configuration was adjusted to the "real" map by rigid transformations which retain the properties of the configuration in Euclidean space.

The results proved to be extremely interesting. Group I (1st year) showed considerably more internal variation than Group II (3rd year), suggesting, as expected, that a learning process had occurred. Group I individuals were probably unfamiliar with large parts of the nearby area, reflecting reduced activity and perceptual overlap.

Direction was also important. As expected, Group II was considerably more accurate than Group I in all directions. Both groups had poor east-west estimates, but Group I was noticibly less accurate. Both showed consistently more accurate estimates north of the campus (away from the CBD) than to the south (toward the CBD). Generally, north distances were under-estimated and south distances over-estimated. Golledge, et al. suggest these biases may be a result of two things, 1) the increased congestion and travel time toward the CBD increase the perceived distances, and 2) the increased density gradient makes distances appear longer. Summing up the distortion between real and perceptual distances, they state:

> (the) relations between perceptual and real distance do tend toward a non-linear relationship as distances become large, but the speed with which non-linearity is approached varies with direction.

It must be remembered, of course, that the distances involved are extremely small so that this observation is only tentative concerning the relation between cognitive and real distance.

Finally, they hypothesized, giving direction to further study, that over-estimation would probably reduce the likelihood of interaction, under-estimation the reverse.

This study is an extremely interesting one in that it uses subjective estimates of distances to derive a perceptual map which can be compared to the "real" map. This comparison was fruitful in defining directional distortions over both groups and between groups. However, the study raises at least one serious question. It does not make clear what type of points were used as the reference points. Specifically, were the stimuli comparable?

The question arises in relation to a more recent study concerned with directional distance distortion in the city. Lee (1970) obtained distance estimates to various well known locations, from a central point on the University of Dundee, both toward and away from the CBD. His results would seem to be in direct conflict with those of Golledge, Briggs, and Demko. Lee found both inward and outward distances overestimated, but the inward distances were less overestimated. His explanation was that people have a focal orientation or bias toward the city centre which is related to "satisfaction" at the centre. This gives rise to a consideration of the problem of value transference. For example, can one compare distances to a church and a pub? Or, does the nature of the destination, that is, its value to the individual, not only affect activity (and hence familiarity), but also the perceived distance? Do places associated with a preferred activity appear nearer?[1] Golledge, Briggs, and Demko left their points undefined. Further studies should examine the problem of value transference in selecting reference points.

Also, it is not clear how the congestion and/or density explanation fits well with activity spaces and familiarity. Golledge, Briggs, and Demko's findings would suggest that individuals are more activity-oriented away from the CBD, at least in Columbus. This, however, seems the reverse of intuitive observations and contrary to Lee's (albeit crude) findings. Lee found some evidence using omission rates, that individuals were more familiar with "in" places than "out" places. Moore (1970) also suggests, on a more theoretical level, the possibility of a "downtown bias".

Nevertheless, this method of obtaining perceptual maps appears extremely useful. Data can be efficiently obtained and the output meaningfully used to examine hypotheses. With information about activity spaces (see, for example, Chapin and Hightower, 1965), it is not inconceivable that a firm link could be established between perceptual distortions of distance, and behaviour.

---

1. See, for example, Bratfisch, (1969) for the connection between emotion and the perception of distance.

VIII.2 Analysis: Place Attributes Applications

In this section applications of multidimensional techniques related to defining place attributes is emphasized. More specifically, this section concentrates on two approaches. First, there is the approach exemplified in the paper by Whipple and Niedell (1970), which, while not specifically spatial, demonstrates an extremely interesting application of MDS, especially for its use with the semantic differential and cluster analysis. Second, there is the work of Demko and Briggs and their development of a spatial choice model. Both MDS and MDU are used, primarily in an inter-city migration context, but it would be applicable in other geographic choice situations.

Whipple and Niedell (1970) were interested in examining black and white perceptions of various department stores in Buffalo, New York. Basically, theirs was a three phase analysis. First, they utilized the semantic differential approach (rating scale using bipolar adjectives or phrases), to collect basic information. Some socioeconomic data were also collected. In all, semantic scales for ten department stores were obtained from 58 respondents, 28 black, in a single integrated neighbourhood. The small number and limited spatial sampling reduce the generality, but the approach can be easily extended.

Preliminary examination of the semantic scores[1] suggested that blacks and whites differed in their shopping behavior and favourableness to the entire set of stores, and interestingly enough, both groups tended not to shop at the store most preferred. It should be pointed out that part of the purpose in this study was to uncover store images so as to develop possible store marketing strategies. Clearly, something more fundamental than favourableness was required.

The second phase consisted of an MDS analysis. A similarities matrix was calculated from the semantic scores using the following formula (p. 6):

$$d_{ij} = \sum_{a=1}^{n} \left| X_{ia} - X_{ja} \right|$$

where d_{ij} = the absolute distance between a pair of stores,
X_{ia} = the semantic score on word (phrase) scale 'a' for store 'i'; $i = 1,\dots,6$.
X_{ja} = the semantic score on word (phrase) scale 'a' for store 'j'; $j = 1,\dots,6$.

1. Initially, they simply compared black and white store rankings via a Spearman rank correlation test, but other methods of analyzing semantic differential scores include factor analysis and profile comparisons.

For each individual, for each pair of stores, the differences on each scale were summed. These individual measures of (dis) similarity were then averaged for each relevant group.

Configurations were then derived for each group (i.e., blacks and whites), with extremely good stress scores in two dimensions ($S_{black} = .0346$; $S_{white} = .0514$). The most striking feature of the two maps were their similarity; black and white perceptions for individual stores were remarkably alike.

The authors then proceeded to ask "why" different stores possessed different images. Using the information from the semantic differential, axes were fitted to the space. The axes seemed to be related to "value" and the "degree of promotion".

It is extremely important to reiterate (V.6) that the axes could not be fitted and interpreted without outside information. The semantic differential is useful in this regard, not only in providing qualitative information on the stimuli, but in providing quantitative information (with the appropriate assumptions) which can be factor analysed. This provides a control both on the number of relevant dimensions and on the stimuli which load highly on a particular dimension, but, unlike the MDS analysis, dimensions derived from a factor analysis, since numerous 'named' variables are involved, are easier to 'label'.

The location of the axes reaffirmed the similarity in race perceptions. It provided information on which stores had similar perceptions (close together) and how they were similar (close together on what dimensions), so that some comment could be made on which stores were and were not competitive. This might have been clarified with a cluster analysis of the stores, that is, by grouping those stores which were perceived similarly.

The third phase involved a cluster analysis[1] of individuals to determine "if significant image differences exist among any subset of the respondents" (p. 9). Clustering individuals maximized store image differences.

Five groups were defined and distance measures between each cluster were computed. The clusters, not the stores, were plotted to determine if different images did exist among groups. With five groups there were (5x4/2) 10 similarities measures. A one dimensional solution resulted (from an MDS analysis), which, since it seemed to vary over education, employment, and income, was termed a "social class" dimension.

---

1. Johnson's (1967) hierarchial, ordinal, cluster technique was used. See section V.6.

Three groups at one end were aggregated into a single class; the remaining two groups formed a second class. Both groups were racially mixed. MDS analyses were performed on each of these two new groups.

Again, results were extremely similar, both to each other and to the black and white configurations earlier. Several individual stores, however, moved significantly. 'Sears', for example, was considered "high quality" by the lower social group, but "intermediate" by the higher social group.

This study is important in demonstrating simply the usefulness of MDS, especially in conjunction with the semantic differential and cluster analysis. Greenberg (undated) has noted the use of the semantic differential to provide the necessary external information. It should be noted, however, that when the semantic scores are used to derive similarities measures, the approach is no longer open-ended. The data result from 'indirect' judgements in which the relevant attributes, the scales of the semantic differential are supplied, a priori. The most important attributes of the stimuli might be excluded, or the attributes included might not be salient to that individual. A strategy in which an MDS configuration is obtained from a "free" method of data collection and then interpreted via information from a more structured technique might prove very successful. If the structured data could not be used to interpret the configuration, one would suspect the structured data to be unrepresentative. This "multiple operationism" approach provides another cross-check on the data.

We might extend the approach exemplified here in other ways. In particular, groups could be defined on a spatial basis and/or stimuli could be specified as to location thereby distinguishing between "branch" stores. Similarly, one might be interested in observing if "value" varied with intervening distance from home to store.

Other physical and social phenomena could be examined. Department stores lend themselves to analysis since they are easily identified, usually well known, and in sufficient number to permit meaningful comparisons. Grocery and other types of stores might also be considered, for both intra-group and inter-group comparisons. Lowry (1970), for example, found cognitive distance varied with nature of the stimulus destination. It would be interesting if low-order-goods stores, such as grocery stores, had a distinct distance dimension, while, say, department stores did not.

More aggregate features such as shopping centres might be considered, but as Downs (1970a) has found, they may be less spatially identifiable. The boundaries of the CBD, for example, are poorly defined physically and mentally.

It would seem possible, however, to use word descriptions, not necessarily of real places, to examine social phenomena, such as residential districts, to provide similarities data. Word descriptions might be supplemented by pictures (see Peterson, 1967), to compare hypothetical neighbourhoods. By systematically altering one expected variable, an "experimental-like" manipulation could be undertaken.

In general, the use of other psychological and sociological techniques to derive similarities data would seem fruitful in studying spatial phenomena.

The second approach we wish to examine is that characterized by the work of Demko and Briggs (1970; 1971) and Demko (1971), concerning spatial choice. They suggest that in many areas of geography, such as shopping, residential site selection, migration, recreation, and acquaintanceship, individuals are engaged in making decisions or choices from a set of possible spatial alternatives based on varying levels of information. A model of such choices necessitates defining a preference function or scale which relates the perceived attributes of an alternative to the relative attractiveness or utility of that alternative, and, a set of rules or mapping, from the subjective preference scale to the actual choices. The mapping rule can be deterministic or probabilistic, and can be weighted to reflect certainty or uncertainty of outcomes.

They conceptualize the choice problem in the following way. For a given situation, there are a large number (n') of alternative locations (i), each characterized by 'j' attributes ($j=1,\ldots,k,\ldots m'$). Therefore, each alternative is defined by a vector (W_{ij}), representing objective physical space. In contrast, each individual (h) ($h=1,\ldots,q$), perceives, or has information on, n alternatives ($n \leq n'$), which defines the effective set of alternatives; the individual may not recognize or value all attributes of the ith alternative. Consequently, each perceived alternative is defined by the vector, ${}_hV_{ij}$ ($i=1,\ldots,n$; $j=1,\ldots m$; $n \leq n'$; $m \leq m'$). Normally, an individual's perceived vector, ${}_hV_{ij}$, is "less than" the corresponding objective vector, W_{ij}, due to the individual's limited mental ability, although, on occasion, with perceptual distortions and additions, ${}_hV_{ij}$ may exceed W_{ij}. The subjective preference <u>scale values</u>, ${}_hR_i$ ($i=1,\ldots,n$), are defined as some composite of the weighted attributes, (${}_ha_j \times {}_hV_{ij}$), where the weight (${}_ha_j$) indicates the relative <u>importance</u> that individual attaches to that attribute. The subjective preference function or mapping rule is that group of functions which relate the scale values, ${}_hR_i$, to the objective vectors, W_{ij}.

It is evident that the choice conceptualization is closely related to the MDS and MDU conceptualizations (sections V, VI). Essentially, there are three problems: 1) defining the m perceived attributes, 2) defining the weighting factors, ${}_h a_j$, for each of the q individuals, and 3) comparing the preference scale values, ${}_h R_i$, to actual choices. Demko and Briggs use MDS and MDU to solve problems 1 and 2, respectively. They do not solve problem 3, although the results of solving the first two problems has certain immediate implications for problem 3. Note that problems 1 and 2 are problems of analysis.

From our earlier discussion, it should be apparent that the perception problem (problem 1), can be solved using similarities data on the n alternatives as input to an MDS algorithm. The output configuration will provide the scale values, ${}_h V_{ij}$, on the m dimensions.

To interpret the dimensions, however, exogenous information is required. Briggs and Demko suggest that, as well as similarities data, ranked data could also be collected on the n stimuli relating to various specific attributes which the experimenter(s) feel are relevant. Factor scores of the exogenous attributes can then be compared to the coordinates of the perception space via a canonical correlation, thereby defining the orientation of the perceptual dimensions and permitting interpretation of the space by the loadings.[1] This approach, Demko and Briggs point out, imposes an orthogonality assumption which may not be valid and difficult to accept if one knows nothing about the interdependencies of the exogenous attributes. However, if one realizes that the MDS analysis requires an orthogonal assumption anyway, the difficulty is less onerous.

As noted above, the configuration provides the scale values, ${}_h V_{ij}$, for the n stimuli. However, by the homogeneity assumption, the subscript h, can be dropped. The scale values are assumed relevant to all individuals. To soften this assumption somewhat, Demko and Briggs suggest that subgroups be analyzed in which perceptions are more or less homogeneous, defined by applying a Q-mode factor analysis to the similarities data. This should reduce within group variations. Another method for grouping is that used by Whipple and Niedell above --- cluster analysis. Either method reduces the possibility of violating the homogeneity assumption.

1. Demko (1971) actually performed this operation. He factored (principle components analysis) the exogenous attributes to derive factor scores for each stimulus on a set of orthogonal dimensions. These factor scores were then compared, by canonical correlation to the scale values derived from the MDS analysis. The high degree of correspondence permitted him to interpret the perceptual dimensions from the interpreted factor dimensions.

The second problem, defining a preference scale, can be solved by means of the MDU algorithm using preferential choice data. The raw data, of course, are simple preference rankings of the stimuli by each individual. The configuration derived, defines the preference scale values, the ${}_hR_i$'s.

It should be noted again, that the subscript h, can be ignored in reference to the weighting factor (${}_ha_j = a_j$), since the weighting of the dimensions is, in effect, a group weighting, not an individual weighting. The subscript h, in the scale values (${}_hR_i$), however, cannot be dropped. Each individual is represented by an ideal point, so that, upon translation to the origin, "relative" individual preference scale values can be determined. One might be tempted to derive individual weights from the equation,

$$V_{ij} = {}_ha_j \times {}_hR_i$$

where ${}_hR_i$ and V_{ij} are known from the perceptual and preference spaces. The ${}_hR_i$, however, are derived from an averaging process of all individuals. Therefore, although individual preference scales can be determined, they are only "relatively" individual and cannot be used to determine an individual's weights (see section VI on the interpretation of the unfolding dimensions). If one wished to examine individual weights, the general or weighted unfolding models might be more appropriate.

The usefulness of developing this type of perception-preference model should be apparent. This approach permits (within the assumptions), identification of the basic dimensions in any choice problem, whether it be migration (Demko, 1971), shopping, recreation, residential site selection, or even plant site selection. These basic dimensions can be related to objective attributes of the stimuli, "to shed light on those attributes which affect spatial decisions" (Demko and Briggs, 1971, p. 52), and to individual attributes such as socio-economic status, location, or activity patterns which may affect the individual's preferences. In this way individual choices may be predictable, a priori.

To summarize, both these approaches (Whipple & Niedell; Demko & Briggs), focus on identifying the relevant attributes of the stimuli considered. Once they have been uncovered, the researcher can examine individual characteristics to see if there are group differences. Whipple and Niedell found socioeconomic differences in perceptions of department stores in Buffalo, but no significant differences in black and white perceptions. Demko and Briggs suggest methods for grouping of individuals which may lead to relations between individual characteristics and perceptions. It is interesting to note, that Demko (1971) found some preference differences, as well as perceptual differences, between groups distinguished by whether they lived in large or small places. However, the focus of attention in Demko and Briggs' work is to identify the objective attributes which affect decisions, not individual characteristics.

VIII.3 Synthesis: Preference Functions

This section is concerned with the use of MDS in the problem of synthesizing preference functions. In particular, the discussion focuses on the work of Rushton (1969a, 1969b, 1969c, 1971) and Ewing (1970), and their efforts to operationalize a choice model for the expenditures of a dispersed population among a set of towns. The data used are actual expenditures of households in a section of Iowa.

Rushton conceptualizes the choice problem in the following way. Households are located different distances from a set of towns varying in their potential opportunities to satisfy the needs of the household. The observed choices made by the household can be considered to be the result of a perceptual process in which all perceived alternatives are ordered as to their relative opportunities and the actual alternative chosen is that which provides the largest expected satisfaction. The object, then, is to recover the tradeoffs involved in this decision process; that is, to uncover how a change in one variable can be compensated for by a change in another variable. This would leave the individual's level of satisfaction, or utility unchanged, making him indifferent to the substitution. These tradeoffs can be determined if a preference function can be defined which orders all possible combinations. In other words, we require an interval scale of preferences.

VIII.4 Procedure

Rushton's analysis proceeds in basically five steps.

The first is to define the stimuli or alternatives. In spatial choices, Rushton emphasizes the need to obtain preference rules which are in no way dependent on the particular density or arrangement of opportunities. For example, we cannot use descriptions of behaviour to derive postulates of spatial behaviour, since the description of behaviour (interaction rates, etc.), are location dependent. The truth of this can be seen by the simple fact that descriptions of behaviour in space in one area are not applicable in another area. Gravity studies of behaviour, for example, derive slope coefficients which vary between areas (eg., Olsson, 1965). Therefore, Rushton distinguishes between descriptions of behaviour in space and postulates of spatial behaviour, which are location independent, but which when applied to a particular area should permit an adequate description of the behaviour in space. He argues that we should be attempting to uncover the rules of spatial behaviour which cannot be done from a description of behaviour in space, but by systematically separating preferences and opportunities.

Consequently, Rushton desires a more general description of choice behaviour divorced from opportunities. Instead of defining his alternatives as single towns, he defines various locational types, made up of combinations of town size and distance. Town size is selected as a surrogate for relevant place attributes, and distance is selected as a surrogate for intervening consumer travel costs. In one study, 30 locational types were distinguished (Rushton, 1969b); in another, 27 locational types were defined (Ewing, 1970).

By using locational types, comparisons are made between types rather than individual towns, which removes the opportunity aspect and makes the data more general. An individual is then considered to choose, or prefer, one locational type over another, rather than one town over another. This also permits a reduced sample size. Often only one or two households would have a choice between two distinct towns. Now, many choices, involving many towns, can be grouped as choices between types.

Having defined the stimuli, the second step is to compute a basic data matrix. One method might be to simply count the number of times a locational type is patronized. Immediately, however, we confuse the problem of opportunity and preference. We can use the frequency method to determine preferences only if the same alternatives are available to all households. Spatially, this would necessitate the rare event of all individuals to be located at a single point. Instead, Rushton distinguishes among the actual choice, choices present but not selected, and choices not present. In this way he makes comparisons only between locational types both of which are available. If only one type is available, there is no information on preferences. If three types are present, then two bits of information are available; the chosen alternative is seen to be preferred to the other two. There is no comparison between the two available but rejected alternatives.

Ewing has modified this approach somewhat. He uses grocery expenditures, in dollar terms, to rank choices. Thus, if three alternatives are available and have been patronized, three bits of information are available. The size of the grocery expenditures permits him to compare the second and third choices.

The third step is to convert the basic data matrix into a probability matrix. This is accomplished using the following formula,

$$P_{AB} = \frac{N(A)}{N(A)+N(B)}$$

where A,B = locational types,
P_{AB} = the probability of choosing A over B,
N(A)= the number of times A is chosen over B when both A and B are available.

If P_{AB} is close to one, then A is strongly preferred to B; if P_{AB} is close to zero, B is strongly preferred to A; and if P_{AB} is close to one-half (.5), A and B are largely indifferent.

The fourth step is to compute distances from the probabilities. Recognizing that probabilities close to one-half indicate that the two alternatives are close together psychologically, "distances" can be defined using the following formula,

$$d_{AB} = |P_{AB} - .5|$$

where d_{AB} = the distance between A and B.
Defining all interpoint distances produces a symmetric matrix which can then be used to obtain a rank order of interpoint distances.

The fifth step is to use the ranked data, interpreted as dissimilarities data, as input in an MDS routine to derive a unidimensional preference scale. This scale has the important feature that it is an interval scale, such that, not simply the order of preference, but the strength of preference for each locational type is determined. By its nature, the scale determines preferences for all combinations of distance and population, including (intermediate) hypothetical combinations of attributes.

VIII.5 Discussion

Before considering the results in more detail, let us reexamine the procedure to discuss several operational and conceptual difficulties.

The first difficulty concerns the definition of alternatives. Rushton arbitrarily assumes that 'distance' and 'population' are useful surrogates for the true decisions variables, and makes no effort to determine if these are in fact relevant to an individual. This is in sharp contrast to much of our earlier discussion, especially the work of Demko and Briggs, whose efforts were directed to identifying the relevant dimensions or variables on which choices were based. In effect, whereas Demko and Briggs adopt a cognitive approach, Rushton adopts a behavioristic approach (see section II). Consequently one cannot help wondering how a different classification would affect the results.

To be fair, however, two points should be made. One, Rushton emphasizes that these variables are not intended to be the relevant variables used in actual decisions. Instead, these are intended as surrogates, or readily obtainable intermediate variables which will act for the more basic variables. In that sense, population and distance have a long tradition in geography. Two, it is perfectly legitimate in logical deductive explanations to begin with initial premises or axioms which are untested, and then to derive their consequences. Of course, the proof of the pudding is in the eating. Here, the proof of the axioms is in the empirical verification of the theorems. Time will dictate the wisdom of his axioms.

A similar criticism, arbitrariness, can be levelled at Rushton's choice of definition of "present but not selected alternatives". How does one decide if an alternative not chosen is available. Rushton uses the figure of 28 miles to determine if a town is available, but offers no reason for this figure. Even if a town is within the prescribed distance, can one be sure the town is perceived as an alternative? There may be a threshold size, for example, below which a town, although near, is not perceived "present" at all.

A second difficulty arises in the necessary assumption that all individuals have similar preference functions. Preference functions can be derived from a single individual making several choices, or several individuals making a single choice. To aggregate the individual choices, which is the only feasible alternative in this instance, necessitates the assumption that the individual preference functions are similar and, therefore, comparable.

Ewing (1970) points out that inter-personal variations could result from three sources; 1) differences in salience weights; that is, differences in the perceived importance of the basic dimensions, 2) differences in perception related to differences in learning, background, and information, and 3) differences in combination rules of the relevant dimensions. The first two problems relate to individual differences in preference and perception which have been discussed in section VII.4, although the model there is more specific than the present discussion.

The third source of interpersonal differences requires some explanation. Gravity models use variables which are combined multiplicatively; multidimensional scaling techniques, in contrast, combine attributes (variables) additively (linearly). In fact, most psychological models combine variables additively (Ewing, 1970), as do most multivariate techniques in general (e.g., multiple regression, factor analysis, discriminant analysis, etc.). Therefore, although the assumption that individual preference functions are similar would seem to be a reasonable and well precedented approach (see discussion in section VI), combining the two variables, distance and population, multiplicatively, to form a single variable, is not well precedented. One might question this approach on these grounds, especially if one notices that the locational types are defined, not over intervals of the population-distance ratio, but over intervals of population <u>and</u> distance.

The question of how the variables should be combined leads to a third question: 'Why use a one-dimensional solution?'. Rushton has anticipated this question since he provides a discussion on a test for intransitivities (Ewing subsequently refined the data by eliminating some interdependent comparisons). Rushton found very few intransitivities, which would suggest that the preferences can be represented by a one-dimensionsl scale. However, his final scale had a (Kruskal) stress value of 43%, which is very high (even Ewing's revised procedure had a stress value of 25%), which suggests that there are numerous intransitivities, and that the preference space is not one-dimensional.[1]

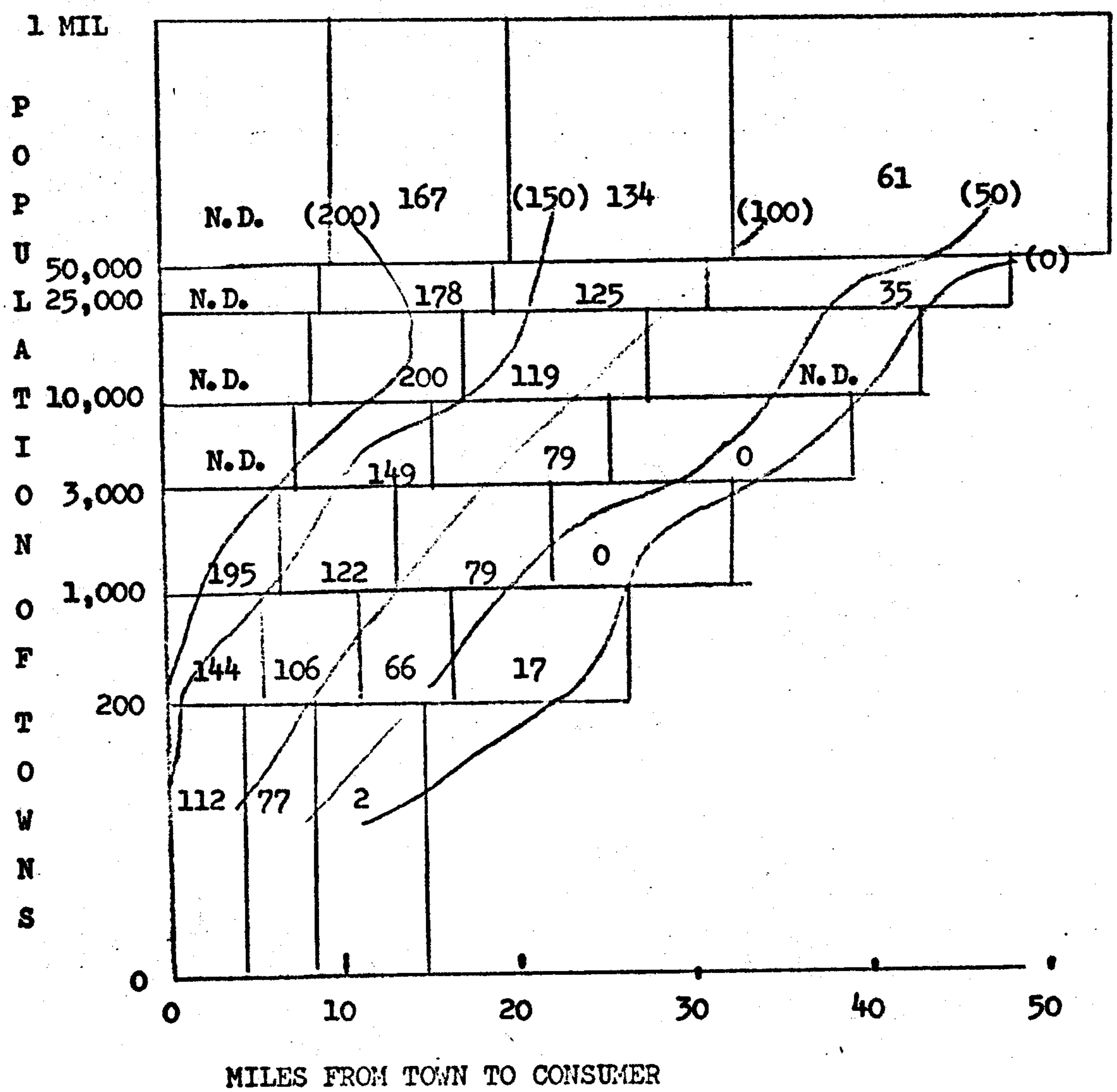
We might well ask how many dimensions should we expect. The similarities data are derived from choices among locational types which are defined by two variables. It would seem reasonable, therefore, to expect a two-dimensional solution. Indeed, it would be interesting to know the stress value for such a configuration. Even if one considers the MDS routine as a technique (which Rushton obviously did), instead of a criterion, the high stress value reduces the confidence in the scale's validity.

---

1. One reason for the discrepancy between the two techniques may be related to Ewing's difficulty in scaling all the locational types (pp. 120-21). He found that if he utilized all types, including those which were very highly preferred (large population and small distance), the scale divided into two groups. At one end were those few highly preferred; and at the other end was a large cluster of all other types. To differentiate the large cluster, he had to eliminate the very highly preferred types. Now, it would seem that if we eliminate these same highly preferred locational types and recalculated the intransitivity measure, the validity of a one-dimensional solution would be more questionable. The reasoning is quite simple. Removing the preferred types reduces the number of possible intransitivities (the denominator of the measure), but probably does not reduce the actual number of intransitivities (the numerator). The intransitivities, seemingly, are concentrated in the middle.

Figure VIII.1

Indifference Surface with Interpolated Locational Types Interval Preference Scores.



Key: Each rectangle represents a single locational type.
Bracketed numbers indicate the degree of preference for that contour.
Unbracketed numbers are scale values derived from one-dimensional MDS solution.

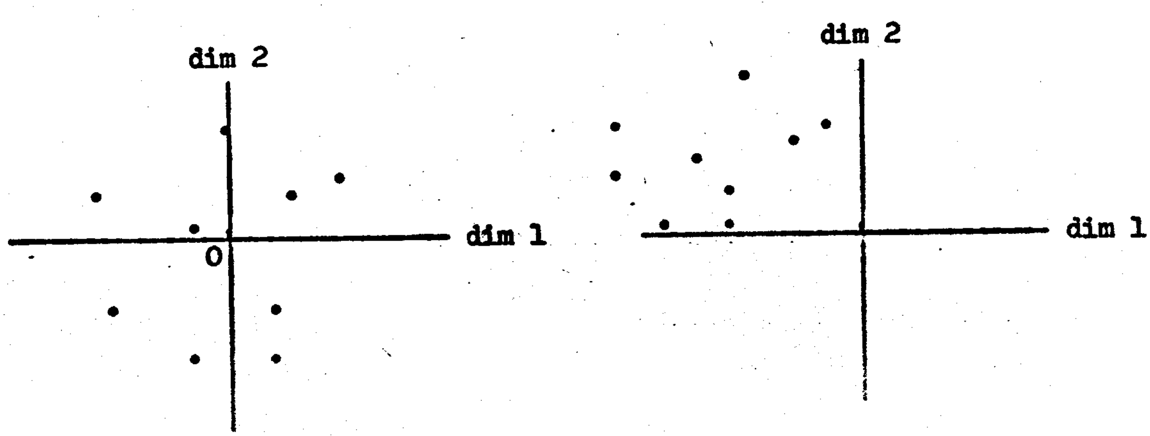
Source: Adapted from Ewing (1970), p. 121, Figure 4.7.

On the other hand, one might ask if a one-dimensional scale is necessary. More specifically: 'Can the tradeoffs between variables be shown in a two-dimensional space?'. Rushton uses the interval scale values to define indifference surfaces as demonstrated in figure VIII.1. The indifference lines are interpolated among the locational types demonstrating the tradeoff between population and distance. It matches our intuitive expectation that preference increases with population, and decreases with distance. In particular, figure VIII.1 shows, for example, that a household would be indifferent to shopping (in Ewing's case, grocery shopping), at a town of about 50,000 population, 30 miles away, and a town of about 1,000 population, 12 miles away. This is indeed a consistent and important result. It provides a quantitative measure of the tradeoff between two familiar geographic variables.

However, can a similar indifference map be constructed, showing the expected tradeoffs, if the preference space solution is two-dimensional? Normally, two-dimensional solutions are portrayed with the origin at the centroid of the points. This need not be the case. The scale is an interval one, so that, assuming Euclidean space, any affine transformation is permitted. Considering that distance is a negative attribute, we can transform the space so that it is portrayed completely in the second quadrant, as in figure VIII.2.

Figure VIII.2

Original and Transformed Stimulus Spaces



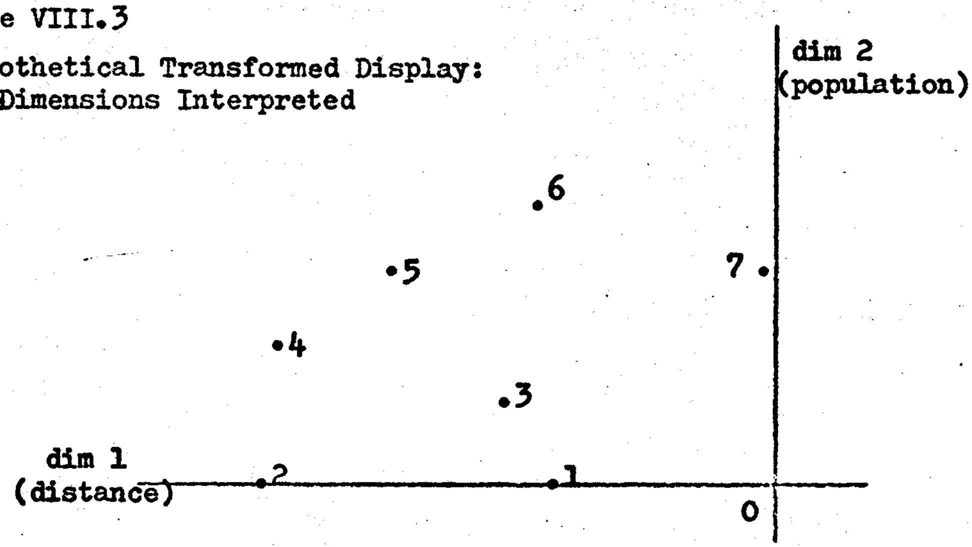
Normal (Original) Display:
0 = centroid of distribution

Transformed Display:
All points of the coordinate form, X = (i,j) where i=0, j=0.

Now consider the hypothetical display shown in figure VIII.3. Assuming that dimension 1 represents distance and dimension 2 represents population (which would need to be interpreted, and may not be the case), points <u>close</u> to the origin will represent relatively <u>high distance values</u> and <u>low population values</u> (i.e., high preference). For example, point 1 would be a locational type which is more distant than locational type 2. Locational type 6 would have a greater population than locational type 3.

Figure VIII.3

Hypothetical Transformed Display: Dimensions Interpreted



However, we desire a single measure of preference. Several methods are possible. One could define preference as the distance from the origin in the transformed space, where distance is defined as,

$$d_k = (d_{k1}^2 + d_{k2}^2)^{1/2} \qquad k=1,\ldots,m$$

where d_k = the distance from point k to the origin,
d_{k1} = the projection of the point k on the first axis,
d_{k2} = the projection of the point k on the second axis,
m = the number of locational types.

Similarly, one could devise other combination rules. The point to recognize is that these new arbitrary preference values result from an additive combination of the variables from a two-dimensional solution. The indifference surface, however, can still be shown with the new preference values attached to the indifference curves. The actual values have no interpretation anyway.

Summarizing, Rushton's approach uses the concepts of 'locational types' and 'available alternatives' to generalize the choice situation. By the method of paired comparisons he is able to derive a distance matrix for locational types, which, when the distances are interpreted as dissimilarity measures permits him, via MDS, to derive an interval scale of preference. This in turn permits an examination of variable tradeoffs by means of an indifference surface.

VIII.6 Extensions

Ewing (1970, Chapter 5) has extended this approach to examine population subgroups. Decisions, he notes, are based on two factors, 1) the "value" of the attributes at the destination (which has been determined in the preference scale), and 2) the characteristics of the individual. Essentially, he is concerned with the assumption that all people attach the same importance to, or have the same weights for, attributes. Or looked at another way, 'Are there distinct subgroups with different preference structures?'.

Since Ewing had information on 12 socioeconomic variables (e.g., education, income, number of members in household, age, etc.), he distinguished dichotomous groups for each variable and tested for differences in preference structures. Only one variable --- number of off-farm workers --- identified groups with significant[1] differences in preference structures. Even when he derived "more basic dimensions" by using a principle components analysis to reduce the 12 original variables to 5 factors, he found no significant differences.

He then reasoned that differences in values should be reflected by differences in spatial preferences. Thus, he hypothesized that households which valued time highly would prefer convenient places. He divided the sample into two groups, distinguishing between those households which made their maximum grocery expenditure at the nearest place and those who did not. In this instance, he found significant differences in preference structures, which could not be explained by differences in opportunities. That is, it was not true that households visiting the nearest place were confined to locations near relatively large places.

In summary, Ewing concluded that preferences do <u>not</u> seem to be correlated to socioeconomic characteristics, but do seem to be related to spatial behaviour. This would suggest, once more, the central role of activity spaces[2] in understanding both perceptions (see Golledge, Briggs, and Demko, 1969; and section VIII.1 above), and motives of individuals.

This idea of activity spaces may be useful in examining Ermuth's (1970) findings, also. Ermuth applied Rushton's approach to a group of households in an urban setting --- Cedar Rapids, Iowa. He divided his sample into two groups on the basis of whether they stated distance was important in their choice of grocery stores, or not.[3] Contrary to what one would expect on the basis of

---

1. Ewing devised a method for determining whether inter-group differences were significant.

2. 'Activity space' is interpreted as a set of nodes frequently visited by an individual. The notion is that activity space is discontinuous. Consequently, two people living in the same area might have different activity spaces if they frequent a different set of nodes. Activity spaces then, are both spatial and hierarchial. The concept might be labelled "activity sectors".

3. He also examined two groups from differant areas of the city.

Ewing's findings, there was no significant difference between the derived preference structures of the two groups. However, Ermuth was concerned with choices made in a relatively small city in which one might expect considerable overlap in activity spaces. Ewing, on the other hand, had little overlap, since few people had the same choices; the population was more dispersed. Breaking Ermuth's sample into mutually exclusive activity spaces might result in significant differences in preferences. Ermuth has noted other differences between the two situations as well. For example, the city shopping situation is much more complex and the distances much smaller in an urban as opposed to a rural setting.

We might end this section by noting several other uses of this approach.

Rushton (1969c) has examined preference structures through time. This is an interesting means of observing the impact of changes in demand (i.e., preference structures), on the urban structure.

Rushton (1971) has examined preference structures across space. He compared preference structures of groups in Iowa and Michigan, at a similar time period, and found considerable agreement.

Rushton (1970) has examined preference structures across shopping activities. He used the scale rankings (derived as above) of three shopping commodities in Iowa and Michigan as input to a further multidimensional analysis. In this case, however, since there is information (the rankings) on each locational type for each of six commodity situations (6x30 matrix) he is able to derive a multidimensional unfolding, joint-space solution. The interesting result (in two-dimensional space), was that five of the six commodity situations were tightly grouped, suggesting that very similar points of view exist across areas and activities. The rogue in the analysis was the point representing the Michigan group's choice of towns for clothing purchases. One suggestion for this divergence was that population may be an inefficient surrogate for clothing opportunities. For example, there may be more clothing stores per population in Iowa than in Michigan. However, the main result was the similarity of spatial preference structures for different shopping activities.

Finally, DeTemple (cited in Golledge and Rushton, 1970), has used this approach to study the diffusion of Harvestore Systems (silos) in part of Iowa. Determining the preference structure for farmers permitted him to obtain the probability of an individual farmer interacting with a given town. These probabilities describe the farmers mean information field (MIF). In regard to the central place hierarchy, he was able to simulate the spread of the innovation quite adequately.

VIII.7 Future Applications

Several modifications and extensions of multidimensional scaling in future geographic work have already been suggested. One area which has been neglected concerns the use of INDSCAL in geography.

INDSCAL tends to focus more on the individual than MDS or MDU. It derives specifically, each individual's perceptual space. Its use in geography, consequently, would seem appropriate in research concentrating on relating individual characteristics (e.g., socioeconomic characteristics), to perceptions and preferences. Whipple and Niedell's study, for example, might have been extended to a careful examination of individuals. Similarly, Ewing's efforts to devise meaningful subgroups might have been approached by comparing individual perceptual spaces (much different data would be required, of course). Cluster analysis might also be used with INDSCAL to group individuals for meaningful analysis. Essentially, INDSCAL permits a very fine stratification of individuals on the basis of perception. Consequently, its use in all sorts of perception problems would seem evident.

In general, multidimensional scaling routines (MDS, INDSCAL, MDU, PREFMAP) would seem to have widespread application in geography. Golledge and Rushton (1970) have suggested at least six areas of possible future application. We now paraphrase their discussion, with some additions and modifications:

1) MDS, MDU, PREFMAP and especially INDSCAL would seem useful in stratifying populations, so that variations in perception and preference can be related to various socioeconomic characteristics.

2) MDS, MDU, PREFMAP, and INDSCAL would seem useful in studying changes in perception and preference over time to: a) better understand the effects of learning and information on behaviour, and b) better anticipate changes in activity patterns which are a result of changes in preference.

3) MDS and MDU would seem useful as a preliminary technique for transforming ordinal data to metric data for use in further analysis.

4) MDS and INDSCAL would seem useful in examining "perceptual gaps" and "distortions" in such areas as store perception, where marketing policy decisions are important.

5) MDS and INDSCAL would be useful in examining perceptual distances and how distortions between the "real" and "perceived" distances affect behavior.

6) MDS, MDU, PREFMAP and INDSCAL would seem useful in redefining such terms as "proximity" and "closeness".

IX Conclusion

Multidimensional scaling techniques have been receiving increasing attention in all areas of perceptual research, including psychology, marketing, political science, and sociology, as well as geography. This trend reflects the applicibility of these techniques to a wide variety of behavioural problems. This trend also reflects the nature of the results, which in many cases match our intuitive expectations, and permit reasonable interpretations.

The few applications which have actually been performed in geography indicate the newness of the approach, rather than any conceptual or operational difficulty. These few examples, however, clearly demonstrate the potential usefulness of the techniques. It is interesting to observe that the distinction between MDS and INDSCAL rests on the assumption of homogeneity of perception, which is the crucial question involved in the "aggregation" problem, briefly examined in section II. It is not inconceivable that these techniques, INDSCAL in particular, may throw some light on this critical area.

Future uses of multidimensional techniques, in general, may in part depend on further technical improvements. One recent development, hierarchial clustering, has increased the flexibility of these techniques and seems to be of increasing importance in present research (see Kruskal, 1972). One unresolved point of difficulty is the problem of determining the proper number of dimensions. Only Kruskal's (1964) rule of thumb exists as a guideline for accepting or rejecting stress (S) values. It seems, however, that researchers are now attempting to define confidence limits for S. If these are established, one of the more subjective aspects of interpretation will be removed.[1]

In conclusion, multidimensional scaling techniques would seem to be applicable in numerous areas of geography --- especially those areas concerned with individual perceptions and preferences and how they affect spatial behaviour.

---

1. Joseph Kruskal discusses some work, published and unpublished, done in this area: "Statistical Aspects of Scaling", Bell-Pen Multidimensional Scaling Workshop, Philadelphia, June 7-10, 1972. From the published work, see for example, Wagenaar, W. A. and Padnos P., "Quantitative Interpretation of Stress in Kruskal's Multidimensional Scaling Technique", Brit. Journ. of Math. Statist. Psychol., 24, pp. 101-10, 1971.

Appendix

An Introduction to the Theory of Measurement

The Representation Problem

The formal definition of a scale relies on an understanding of relational systems. Our problem is to assign numbers to stimuli in such a way that all relations will be represented by the numbers. This involves, on the one hand, a real world system of stimuli, and on the other hand, an abstract numerical system. Each is a relational system.

Suppes and Zinnes make the concept more precise (1963, p. 5):

> A relational system is a finite sequence of the form $U = \langle A, R_1, \ldots, R_n \rangle$ where A is a nonempty set of elements called the domain of the relational system, U, and $R_1, \ldots, R_n$ are relations on A.

The pointed brackets signify a relational system. A simple case might be the situation where A is the set of grocery stores in a city and R is the relation "is preferred to". Symbolically this might be represented by $\succ$. This defines a strong order relation. A weak order relation, "is at least as preferred to", might be symbolically represented by $\succsim$, and includes the possibility of indifference. Notice that the symbol R is no more or less appropriate than the symbols $\succ$, or $\succsim$. The latter are merely introduced as being more familiar.

Another simple example might be the situation, $U = \langle B, D \rangle$, where B is the set of, say, towns in an area, and D is the relation representing individual judgements between the similarities of pairs of towns, i.e., for any a, b, c, d, in B, ab D cd, if and only if a is more similar to b than c is to d. D is an order relation, but involving two dyads or couples rather than a single pair. The latter is a binary relation whereas the former is a quadratic relation.

The connection between the empirical system and the numerical system is is the key to scaling. The connection is the concept of isomorphism or "same structure". Following Suppes and Zinnes (1963, p. 6), consider two similar relational systems[1], $U = \langle A, R \rangle$ and $Z = \langle B, S \rangle$. Then these two systems are isomorphic images if there is a one to one function, f, mapping A onto B such that for every a and b in A,

$$aRb \text{ if and only if } f(a)\ S\ f(b)$$

This may be generalized to more than binary relations (see Suppes and Zinnes, 1963, p. 6).

1. "Similar" is analogous to "order" in matrix algebra. Two systems must have the same number and type of relations (i.e., binary and quadratic are two types of relations).

The representation problem then can be clearly defined as determining an isomorphic[1] mapping from the empirical relational system to the abstract numerical system.

This terminology also allows us to formally define a scale. Consider U to be an empirical relational system, Z to be a full (i.e., the domain is the set of real numbers) numerical relational system, and f to be a function which homomorphically (see footnote 1 below) maps U onto a portion of Z. Then the ordered triple $\langle U,Z,f \rangle$ is defined as a scale (Suppes and Zinnes, 1963, p. 11).

The Uniqueness Problem

The uniqueness problem is concerned with how much freedom there is in constructing the mapping function, f, which generates the scale. Different measurements provide different amounts of information. As the amount of information provided about the order relations of the empirical system increases, the amount of freedom in constructing a representative numerical system declines.

In general, several classes of functions are recognized. Each varies in the amount of information contained, i.e., the degree of uniqueness. Isard (1969, pp. 172-176) defines at least five classes of functions: ordinal, lineal, scalinal, transinal, and cardinal.

Ordinal

An ordinal measurement or scale is a simple ranking of the objects, for example, x is preferred to y, y to z, and x to z. A numerical function which represents this system might be, $f(x) = 3$, $f(y) = 2$, and $f(z) = 1$. This system replicates all the order relations, i.e., $f(x) > f(y) > f(z)$. However, this numerical system is not the only one possible. We could let $f(x) = 99$, $f(y) = 57$, and $f(z) = 14$. Order would still be preserved. In fact, there are an infinite number of possibilities. Thus, an ordinal ranking is said to be unique up to an order preserving (positive monotone) transformation.

It is apparent that the ordinal constraint is not a stringent one. However, in some cases, empirical systems cannot be replicated even by an ordinal function. This occurs when judgements are intransitive, that is, when $f(x) > f(y)$, $f(y) > f(z)$, but $f(z) > f(x)$. Obviously, this situation is illogical, but it does occur.

One explanation, put forward by L. L. Thurstone (1927) and known as the Law of Comparative Judgement, is that each stimulus fluctuates randomly about

1. If two stimuli in the domain are assigned the same number, the assignment is less constrictive but normally considered still valid. This is termed "homomorphic".

some modal value, so that at any given time the actual value ellicited may not be the "true" value. In this way single rankings can be intransitive although repeated trials should clarify the true ordering.

Thurstone's conception introduces in effect a probability element. Consequently, as well as the pure conception of transitivity which allows for no error, three types of stochastic transitivity are recognized: strong, moderate, and weak. The number of times that x is judged greater than y can be expressed as a probability. For example, if out of 10 replications, x is judged greater than y, 7 times, then the probability of $x > y$ is .7, i.e., $p(x>y) = .7$. Equality would occur when $p(x>y) = .5$. <u>Strong stochastic transitivity</u> then can be defined as,

$$p(x>y) \geqslant .5, \quad p(y>z) \geqslant .5 \Longrightarrow p(x>z) \geqslant \text{maximum}(\ p(x>y),\ p(y>z))$$

<u>Moderate stochastic transitivity</u> is defined as,

$$p(x>y) \geqslant .5, \quad p(y>z) \geqslant .5 \Longrightarrow p(x>z) \quad \text{minimum}(\ p(x>y),\ p(y>z))$$

<u>Weak stochastic transitivity</u> is defined as,

$$p(x>y) \geqslant .5, \ p(y>z) \geqslant .5 \Longrightarrow p(x>z) \geqslant .5$$ (Coombs, 1964, p. 106)

The distinction between types of stochastic transitivity is important when measurement models are used to test hypotheses. For example, we might ask, "Can the preferences of ten individuals be represented by a unidimensional scale?". Then, if a scale could be determined which satisfied the requirements of strong stochastic transitivity, rather than weak stochastic transitivity, the hypothesis would be strongly supported. The importance of the concept of stochastic, rather than deterministic, transitivity is that it provides an immediate method of resolving conflicts or "errors" in the data. The hypothesis is not rejected on the basis of a few "random" intransitivities. Instead, probabilities are determined, which, if they meet at least the conditions of weak stochastic transitivity, immediately determine the ordering of the objects.

<u>Lineal</u>

Consider the situation where not only the order of stimuli is provided, but also the order of differences, that is, not only that, $f(x) > f(y) > f(z)$, but also, for example, that $f(x-z) > f(x-y) > f(y-z)$. This confines the choice of a function much more than the simple ordinal scale. For instance, the previous ordinal function, where $f(x) = 3$, $f(y) = 2$, and $f(z) = 1$, is now insufficient, since it does not account for all the order relations. One function which would satisfy these relations is, $f(x) = 4$, $f(y) = 2$, $f(z) = 1$. Many others are possible. Now, however, a simple order preserving transformation is not admissable. To preserve the order of the intervals, a positive linear transformation (also termed "increasing linear"), is required. Operationally, this involves multiplying each value by a positive constant and then adding another constant:

$$f(x) = ax + b \qquad a > 0,\ b > 0$$

A lineal scale then, unique up to a positive linear transformation, uniquely determines all values except for an origin and a unit of measurement. And since it deals not only with the order of values but also with differences or intervals between values, it is termed an interval scale.

There are, however, several different types of interval scales. The classic example of an interval scale is that of temperature; another is that of intelligence quotient (IQ). In both these cases the size of the interval, given the scale can be determined. For example, object A might be 5.4 Centigrade degrees warmer than object B; or individual X may have an IQ 5 points higher than individual Y. The intervals are precisely calibrated. If, on the other hand, only the order of the differences is known, a type of interval scale known as an ordered metric is generated. These ordered metrics, represent a whole family of ordered interval scales. In our original example, we ordered all intervals. Other ordered metrics might only order successive intervals (i.e., between adjacent stimuli). The number of possible intervals which can be ordered for n objects is $n(n-1)/2$. There are, of course, numerous subsets. The concept of ordered metrics is crucial to nonmetric multidimensional techniques.[1]

Scalinal

Consider now the situation where an individual is able to provide information not only on the relative differences, but also on the magnitude of the stimuli directly. For example, he perceives x to be twice as much as y. This is a scalinal function. It replicates all the order relations present in a lineal function but also provides a fixed reference point. Clearly, it is meaningless to say that x is three times as great as y, unless there is a third object or reference point from which to determine the first two. For example, there might be a reference point from which x would only be twice as great as y.

Since a scalinal scale has a fixed reference point, it is a type of ratio

1. For a fuller discussion of "higher order metrics", see Fishburn P. Decision and Value Theory, New York: John Wiley & Sons, Inc., 1964.

scale. Consequently, it is said to be unique up to a positive scalar transformation, or, another common term, unique up to a proportionality transformation. Both terms simply mean multiplication by a positive constant is permited:

$$f(x) = ax \qquad a > 0$$

Notice all such transformations leave the zero point unchanged.

It is interesting to note that many common measurements involve ratio scales. Weight and distance are two such examples. Each has a natural zero point and arbitrary unit of measure. For example, the definition of a kilometre is no less arbitrary than the definition of a mile.

Transinal

A transinal function records not just relative differences but absolute differences. The size of the interval is absolutely determined. Consequently, a system which is defined by a function where $f(x) = 3$, $f(y) = 2$, $f(z) = 1$, measured on a transinal scale could be equally represented by a function where, $f(x) = 9$, $f(y) = 8$, and $f(z) = 7$. The absolute differences would be unchanged. This type of measurement is unique up to the addition or subtraction of a constant:

$$f(x) = x \pm b \qquad b \neq 0$$

Notice this presupposes an absolute unit of measure but does not specify a zero.

Cardinal

Finally, a cardinal measure is simply the combination of the features of the scalinal and transinal functions. Both absolute magnitudes and differences are defined. Consequently, this type of scale is also termed an absolute scale and allows for no transformations.

Meaningfulness

Closely related to the two basic problems of scaling, the representation and uniqueness problems, is what Suppes and Zinnes (1963) have termed the meaningfulness problem. The inferences and conclusions drawn from the numerical solution must be meaningful in relation to the amount of information input. For example, some statements are not testable with ordinal data. Thus, for a statement to be truthful and meaningful, its truth or falsity must be invariant across admissible transformations (Suppes and Zinnes, 1963, p. 66). In this way, the meaningfulness problem is closely tied to the uniqueness problem.

References

Arrow, Kenneth J. (1967) "Public and Private Values", in Sidney Hook, (ed), Human Values and Economic Policy, New York: New York University Press, pp. 3-21.

Beals, R., Krantz, D. H. & Tversky, A. (1968) "Foundations of Multidimensional Scaling", Psychological Review, Vol. 75, pp. 127-142.

Bennett, J. F. & Hays, W. L. (1960) "Multidimensional Unfolding: Determining the Dimensionality of Ranked Preference Data", Psychometrica, Vol. 25, pp. 27-43.

Berry, B. J. L. (1964) "Approaches to Regional Analysis: A Synthesis", Annals of the Association of American Geographers, Vol. 54, pp. 2-11.

Bratfisch, Oswald (1969) "A Further Study of the Relation Between Subjective Distance and Emotional Involvement", Acta Psychologica, Vol. 29, pp. 244-55.

Briggs, Ronald (1969) The Scaling of Preferences for Spatial Locations: An Example Using Shopping Centres, Columbus, Ohio: M.A. Thesis, Dept. of Geography, Ohio State University. (Cited in Golledge & Rushton, 1970)

Briggs, Ronald (1972) "Urban Cognitive Distance", unpublished paper, University of Texas at Austin.

Carroll, J. Douglas (1971) "An Overview of Multidimensional Scaling Methods Emphasizing Recently Developed Models for Handling Individual Differences", in Charles King & Douglas Tigert, (eds.), Attitude Research Reaches New Heights, Chicago: American Marketing Assoc., pp. 235-262.

Chapin, Stuart F. & Hightower, Henry C. (1965) "Household Activity Patterns and Land Use", Journal of the American Institute of Planners, pp. 222-231.

Chaplin, James P. & Krawiec, T. S. (1968) Systems and Theories in Psychology, New York: Holt, Rinehart & Winston, Inc., 2nd Edition.

Christaller, Walter (1933, 1966) Central Places in Southern Germany, (Originally published as: Die Zentralen Orte in Suddeutschland) Englewood Cliffs, N. J.: Prentice-Hall, Inc.

Coombs, Clyde H. (1964) A Theory of Data, New York: John Wiley and Sons, Inc.

Coombs, C. H., Dawes, R. M. & Tversky, A. (1970) Mathematical Psychology: An Elementary Introduction, Englewood Cliffs, N. J.: Prentice-Hall, Inc.

Coxon, A. P. M. (1972) "Differential Cognition and Evaluation: An Introduction to Carroll and Chang's Multidimensional Scaling Models", Research Memoranda, No. 1, Dept. of Sociology, University of Edinborough.

Cross, D. V. (1965) Metric Properties of Multidimensional Stimulus Control, unpublished Ph.D. thesis, University of Michigan. (Cited in Demko & Briggs, 1971; and Coombs, Dawes & Tversky, 1970.)

Demko, Donald (1971) Perception and Preference Structures for Urban Areas in Southern Ontario, Paper Prepared for the Canadian Council on Urban and Regional Research.

Demko, D. & Briggs, R. (1970) "An Initial Conceptualization and Operationalization of Spatial Choice Behavior: A Migration Example Using Multi-dimensional Unfolding", Proceedings of the Canadian Assoc. of Geog., pp. 79-86.

Demko, D. & Briggs, R. (1971) "A Model of Spatial Choice and Related Operational Considerations", Proceedings of the American Assoc. of Geog., Vol. 3, pp. 49-52.

Downs, Roger (1970a) "The Cognitive Structure of an Urban Shopping Centre", Environment and Behavior, Vol. 2, pp. 13-39.

Downs, Roger (1970b) "Geographic Space Perception: Past Approaches and Future Prospects", in C. Board, R. Chorley, P. Haggett & D. Stoddart, (eds.), Progress in Geography: International Reviews of Current Research, London: Edward Arnold.

Ermuth, Frederick (1970) "Intra-Urban Space Preferences for Grocery Shopping", unpublished manuscript, Dept. of Geog., University of Iowa.

Ewing, Gordon (1970) An Analysis of Space Preference Structures, unpublished Ph.D. thesis, Dept. of Geog., McMaster University.

Flanigan, F. J. & Kazdan, J. L. (1971) Calculus Two: Linear and Nonlinear Functions, Englewood Cliffs, N. J.: Prentice-Hall, Inc.

Georgescu-Roegen, Nicholas (1971) The Entropy Law and Economic Process, Cambridge, Mass.: Harvard University Press.

Golledge, Reginald G. (1969) "The Geographical Relevance of Some Learning Theories", in K. Cox & R. Golledge, (eds.), Behavioral Problems in Geography: A Symposium, Evanston Ill.: Northwestern University.

Golledge, R. G., Briggs, R. & Demko, D. (1969) "The Configuration of Distances in Intra-Urban Space", Proceedings of the Assoc. of Am. Geog., Vol. 1, pp. 60-65.

Golledge, R. G. & Rushton, Gerard (1970) "Multidimensional Scaling: Review and Geographical Applications", Paper Prepared for the International Geographical Union, Commission on Quantitative Geography, A. Mickiewicz University, Poznan, Sept. 1970.

Gould, P. R. (1967) "Structuring Information on Socio-Temporal Preferences", Journal of Regional Science, Vol. 7, pp. 259-274.

Gould, Peter (1969) "Problems of Space Preference Measures and Relationships", Geographical Analysis, Vol. 1, pp. 31-44.

Green, Paul & Carmone, Frank (1968) "Multidimensional Scaling: An Introduction and Comparison of Nonmetric Unfolding Techniques", (mimeo, Oct. 1968). Also published, Journal of Marketing Research, Vol. 6, pp. 330-341, 1969.

Green, Paul & Carmone, Frank (1971) "Stimulus Context and Task Effects on Individuals' Similarities Judgements", in Charles King & Douglas Tigert, (eds.) Attitude Research Reaches New Heights, Chicago: American Marketing Assoc., pp. 263-299.

Greenberg, M. G. (1969) "A Variety of Approaches to Non-Metric Multidimensional Scaling", Paper Presented at 16th International Meeting of the Institute Management Sciences, N. Y. 1969.

Greenberg, M. G. (undated) "Some Applications of Nonmetric Multidimensional Scaling", (mimeo).

Guttman, L. (1968) "A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration", Psychometrika, Vol. 33, pp. 469-508.

Harvey, David (1969) Explanation in Geography, London: Edward Arnold.

Huff, D. L. (1960) "A Topographic Model of Consumer Space Preferences", Papers and Proceedings of the Reg. Sc. Assoc., Vol. 7, pp. 19-28.

Isard, Walter, Smith, Tony, et al. (1969) General Theory: Social, Political, Economic and Regional, Cambridge: MIT Press.

Johnson, S. C. (1967) "Hierarchial Clustering Schemes", Psychometrika, Vol. 32, pp. 241-254.

Kruskal, J. B. (1964a) "Multidimensional Scaling by Optimizing Goodness of Fit", Psychometrika, Vol. 29, pp. 1-27.

Kruskal, J. B. (1964b) "Nonmetric Multidimensional Scaling: A Numerical Method", Psychometrica, Vol. 29, pp. 115-129.

Kruskal, J. B. (1972) "Linear Transformation of Multivariate Data to Reveal Clustering", in A. K. Romney, R. N. Shepard & S. B. Nerlove, (eds.), Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, Vol. 1, New York: Seminar Press.

Lee, Terence (1970) "Perceived Distance as a Function of Direction in the City", Environment and Behavior, Vol. 2, No. 1.

Lingoes, J. C. & Roskam, E. (1970) "An Empirical Study of Two Multidimensional Scaling Algorithms", (mimeo).

Loesch, August (1954) The Economics of Location, New Haven: Yale University Press.

Lowry, Robert A. (1970) "Distance Concepts of Urban Residents", Environment and Behavior, Vol. 2, pp. 52-73.

McGee, V. (1966) "The Multidimensional Analysis of 'Elastic' Distances", The Brit. Journ. of Math. and Stat. Psychology, Vol. 19, pp. 181-195.

McGee, V. (1968) "The Multidimensional Scaling of N-Sets of Similarity Measures: A Non-metric Individual Differences Approach", Multivariate Behavioral Research, Vol. 2, pp. 233-248.

Moore, Eric G. (1970) "Some Spatial Properties of Urban Contact Fields", Geog. Analysis, pp. 377-383.

Nagel, Ernest (1961) The Structure of Science, New York: Harcourt, Brace & World, Inc.

Neidell, Lester A. (1969) "The Use of Nonmetric Multidimensional Scaling in Market Analysis", Journ. of Marketing, Vol. 33, pp. 37-43.

Olsson, Gunnar (1965) Distance and Human Interaction: A Review and Bibliography, Philadelphia: Regional Science Research Institute.

Olsson, Gunnar (1969) "Inference Problems in Locational Analysis", in K. Cox & R. G. Golledge, (eds.), Behavioral Problems in Geography: A Symposium, Evanson Ill.: Northwestern University.

Peterson, George L. (1967) "A Model of Preference: Quantitative Analysis of the Perception of the Visual Appearance of Residential Neighbourhoods", Journ. of Reg. Sc., Vol. 7.

Proshansky, H., Ittelson, W. & Rivlin, L. (1970) "Freedom of Choice and Behavior in a Physical Setting", in Proshansky, H., Ittelson, W. & Rivlin, L. (eds.), Environmental Psychology: Man and His Physical Setting, New York: Holt, Rinehart & Winston, Inc., pp. 173-183.

Rushton, Gerard (1969a) "Analysis of Spatial Behavior by Revealed Space Preference", Annals of the Assoc. of Am. Geog., Vol. 59, pp. 391-400.

Rushton, G. (1969b) "The Scaling of Locational Preferences", in K. Cox & R. G. Golledge, (eds.), Behavioral Problems in Geography: A Symposium, Evanston Ill.: Northwestern University, pp. 197-227.

Rushton, G. (1969c) "Temporal Changes in Space Preference Structures", Proceedings of the Assoc. of Am. Geog., Vol. 1, pp. 129-132.

Rushton, G. (1970) "Joint Space Analysis of Space Preference Structures", unpublished paper, Dept. of Geography and Institute of Urban and Regional Research, The University of Iowa.

Rushton, G. (1971) "Preference and Choice in Different Environments", Proceedings of the Assoc. of Am. Geog., Vol. 3, pp. 146-150.

Shepard, R. N. (1962) "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function", Part I, Psychometrika, Vol. 27, pp. 125-139; Part II, Psychometrika, Vol. 27, pp. 219-246.

Shepard R. N. (1966) "Metric Structures in Ordinal Data", Journal of Math. Psychology, Vol. 3, pp. 287-315.

Simon, H. A. (1957) Models of Man, New York: John Wiley & Sons, Inc.

Stefflre, V. (1971) "Some Applications of Multidimensional Scaling to Social Science Problems", in Charles King & Douglas Tigert, (eds.), Attitude Research Reaches New Heights, Chicago: American Marketing Assoc., pp. 235-62.

Stokes, Donald E. (1971) "Cross-Level Inference as a Game Against Nature", in J. L. Bernd, (ed.), Mathematical Applications in Political Science, IV, Charlottsville, Va.: The University Press of Va., pp. 62-83.

Suppes, Patrick & Zinnes, Joseph L. (1963) "Basic Measurement Theory", in R. D. Luce, R. R. Bush & E. Golanter, (eds.), Handbook of Mathematical Psychology, Vol. 1, New York: John Wiley & Sons, Inc., pp. 1-76.

Tobler, W. R. (1967) "An Empirical Map Projection of the United States", unpublished paper, Ann Arbor, Michigan: Dept. of Geog., University of Michigan.

Tobler, W., Miekle, H., & Detwyler, T. (1970) "Geobotanical Distance Between New Zealand and Neighbouring Islands", Bioscience, pp. 537-542.

Torgerson, W. S. (1958) Theory and Methods of Scaling, New York: J. Wiley & Sons.

Wagenaar, W. A. & Padnos, P. (1971) "Quantitative Interpretation of Stress in Kruskal's Multidimensional Scaling Technique", Brit. Journ. of Math. Statist. Psychol., Vol. 24, pp. 101-10.

Whipple, Thomas W. & Neidell, Lester A. (1970) "Black and White Perceptions of Competing Stores: An Application of Nonmetric Multidimensional Scaling", unpublished paper, Dept. of Marketing, University of Buffalo.

Wish, Myron (1971) "Individual Differences in Perceptions and Preferences Among Nations", in Charles King & Douglas Tigert, (eds.), Attitude Research Reaches New Heights, Chicago: American Marketing Assoc.

Young, Forrest W. (1968) A Fortran IV Program for Nonmetric Multidimensional Scaling, Chapel Hill, N. C.: Report No. 56, The L. L. Thurstone Psychometric Laboratory, University of North Carolina.